

Datakvalitet i Transportstyrelsens register över vägtrafikolyckor (Strada)

En beskrivning ur ett användarperspektiv

Christian Howard
Åsa Forsman

vti

VTI rapport 1261
Utgivningsår 2026
vti.se/publikationer

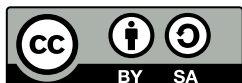
VTI rapport 1261

Datakvalitet i Transportstyrelsens register över vägtrafikolyckor (Strada)

En beskrivning ur ett användarperspektiv

Christian Howard

Åsa Forsman



Om inget annat anges är publikationen licensierad enligt [CC BY-SA 4.0](https://creativecommons.org/licenses/by-sa/4.0/), omslaget omfattas inte av licensen./Unless otherwise stated, the publication is licensed under [CC BY-SA 4.0](https://creativecommons.org/licenses/by-sa/4.0/), the cover is not included in the license.

Författare: Christian Howard (VTI), Åsa Forsman (VTI)

Diarienummer: 2024/0474-8.3

Publikation: VTI rapport 1261

Utgiven av VTI 2026

Publikationsuppgifter – Publication Information

Titel/Title

Datakvalitet i Transportstyrelsens register över vägtrafikolyckor (Strada). En beskrivning ur ett användarperspektiv / Data quality in the Swedish Transport Agency's register of road traffic crashes (Strada). A description from a user perspective

Författare/Author

Christian Howard (VTI, orcid.org/0000-0003-0969-2771)

Åsa Forsman (VTI, orcid.org/0000-0003-4680-4795)

Utgivare/Publisher

VTI, Statens väg- och transportforskningsinstitut/
Swedish National Road and Transport Research Institute (VTI)
www.vti.se/

Serie och nr/Publication No.

VTI rapport 1261

Utgivningsår/Published

2026

VTI:s diarienum/Reg. No., VTI

2024/0474-8.3

ISSN

0347-6030

DOI

<https://doi.org/10.65151/vti680231>

Projektnamn/Project

Utveckling av olycksdatabasen Strada/Development of the crash database Strada

Uppdragsgivare/Commissioned by

Slutrapporten är framtagen med ekonomiskt stöd från Skyltfonden, Trafikverket/This final report is produced with financial support from Skyltfonden, Swedish Transport Administration

Kort sammanfattning

Strada är ett nationellt register för vägtrafikolyckor som kombinerar data från både polisen och akutsjukvården. Målet är att tillhandahålla detaljerad information om både olycksomständigheter och skador för forskning, utvärdering och måluppföljning. Med Stradas centrala roll i svenskt trafiksäkerhetsarbete är hög datakvalitet av stor vikt.

Syftet med denna studie är att beskriva Stradas datakvalitet ur ett användarperspektiv och att ta fram rekommendationer för att förbättra datakvalitet och användarbarhet. Detta görs genom att utforska frågeställningar kring partiellt bortfall, matchning av rapporter, datakopplingar till nationella vägdatan (NVDB), samt behovet av dokumentation och metadata. Metodansatsen innehåller både kvantitativa och kvalitativa delar och baseras på analys av Stradadata, studie av dokument, och inhämtning av expertinformation.

Resultaten visar att täckningen gällande dödade inte är fullgod, användarbarheten av kopplade NVDB-data är låg, och att det finns ett behov av bättre dokumentation. Sett till partiellt bortfall indikerar resultaten en förbättring för polisrapporter och en försämring för sjukvårdsrapporter över tid. Vidare har en beskrivning av matchningsalgoritmen tagits fram. Utifrån studiens resultat rekommenderas Transportstyrelsen att:

1. föra dialog med departement och berörda myndigheter om polisens nedprioritering av rapportering till Strada
2. prioritera inkludering av information om personer med skyddade personuppgifter
3. utöka värdemängderna för Stradas variabler för tydligare indelning av saknade värden
4. säkerställa möjlig koppling till polisens interna utredningsmaterial
5. kartlägga och dokumentera matchningsalgoritmen i detalj
6. utveckla en användarvägledning som beskriver olyckskopplingen till NVDB och hur kopplingen kan användas
7. återinföra vägtyp som polisrapporterad variabel med samma värdemängd som i NVDB
8. utveckla polisens och sjukvårdens rapporteringsverktyg så att endast relevanta väglänkar kopplas från NVDB och att dessa struktureras i separata variabler
9. dela upp nuvarande användarhandledning för Strada uttagswebb i två dokument med registerinformation och uttagsinformation
10. skapa och bifoga en maskinläsbar variabelmetadatabil till varje datauttag, samt att
11. erbjuda data i textformat som alternativ till Excelformat.

Nyckelord

Trafiksäkerhet, vägtrafik, olycksdata, skadedata, Strada

Abstract

Strada is a national registry for road traffic crashes that combines data from the police and emergency hospitals. The aim is to provide detailed information on both crash circumstances and injuries for research, evaluation, and target monitoring.

The purpose of this study is to describe Strada's data quality from a user perspective and to propose recommendations for improving quality and usability. Partial missing data, report matching, data linkages to the national road database (NVDB), documentation and metadata are explored. The study is based on analysis of Strada data, document studies, and the collection of expert knowledge.

The results show that the coverage of fatalities is insufficient, the usability of linked NVDB data is low, and documentation needs improvement. Regarding partial missing data, the results indicate an improvement over time for police reports and a decline for healthcare reports. Furthermore, a description of the matching algorithm has been developed. Based on this description and the results, it is recommended to

- 1) hold discussions with the ministry and the relevant authorities concerning police deprioritisation of reporting to Strada
- 2) prioritise inclusion of individuals with protected personal data
- 3) expand the value sets of variables to provide clearer categorisation of missing values
- 4) ensure possible linkage to the police's internal investigation materials
- 5) map the matching algorithm in detail
- 6) develop a user guide describing the linkage to NVDB
- 7) reintroduce road type as a police-reported variable with the NVDB value set
- 8) develop the reporting tools to only include relevant road links from NVDB and to structure them into separate variables
- 9) divide the user manual into two documents with registry information and extraction information
- 10) create a machine-readable variable metadata file to be included with extracted data, and
- 11) offer data in text format as an alternative to the Excel format.

Keywords

Traffic safety, road traffic, crash data, injury data, Strada

Innehållsförteckning

Publikationsuppgifter – Publication Information	5
Kort sammanfattning.....	6
Abstract.....	7
Sammanfattning	9
Summary	11
Förord.....	13
1. Introduktion.....	14
2. Syfte	16
3. Metod.....	17
3.1. Stradadata.....	17
3.2. Dokument.....	17
3.3. Expertinformation	18
4. Resultat.....	19
4.1. Dödade enligt polis och sjukvård.....	19
4.2. Polisrapporterade data.....	22
4.3. Sjukvårdsrapporterade data.....	23
4.4. Matchning av rapporter.....	25
4.4.1. Matchning på personnivå.....	26
4.4.2. Matchning på olycksnivå, endast sjukvårdsrapporter	27
4.4.3. Matchning på olycksnivå, polis- och sjukvårdsrapporter	27
4.4.4. Resultat av matchning i Strada år 2022 och 2023.....	27
4.5. Data från Nationella vägdatabasen (NVDB).....	33
4.6. Dokumentation och metadata.....	35
4.6.1. Övergripande dokumentation.....	36
4.6.2. Tabellstruktur för datauttag.....	36
4.6.3. Variabelspecifik metadata.....	37
5. Diskussion	39
6. Slutsatser och rekommendationer	42
Referenser	43
Bilaga 1: Andel saknade värden polis.....	44
Bilaga 2: Andel saknade värden sjukvård	46

Sammanfattning

Strada (Swedish traffic accident data acquisition) är ett nationellt register för vägtrafikolyckor som kombinerar data från både polisen och samtliga akutsjukhus som ger akutsjukvård inom ortopedi eller kirurgi i Sverige. Genom att kombinera dessa datakällor är målet att tillhandahålla detaljerad information om både olycksomständigheter och skador för forskning, utvärdering och måluppföljning. Data från Strada är en grundpelare för förståelsen av trafiksäkerhetsutvecklingen i Sverige och är viktiga för att bedöma effekten av trafiksäkerhetsåtgärder. Utöver polis- och sjukvårdsrapporter inkluderar registret också kopplade data från Nationella vägdatabasen (NVDB) och Vägtrafikregistret. Det är viktigt att granska Stradas data och dokumentation ur ett användarperspektiv, så att välgrundade avvägningar kan göras kring eventuella brister och biaser i data.

Syftet med denna studie är att beskriva Stradas datakvalitet ur ett användarperspektiv och att ta fram rekommendationer för att förbättra Stradas datakvalitet och användbarhet. Detta genom att utforska följande frågeställningar:

- Vilken kvalitet har data som rapporteras in från polisen och hur har detta förändrats över tid?
- Vilken kvalitet har data som rapporteras in från sjukvården och hur har detta förändrats över tid?
- Hur fungerar matchningen av polis- och sjukvårdsrapporter?
- Hur fungerar kopplingen mot NVDB och vilken information tillförs Strada genom denna koppling?
- Vilken dokumentation och vilka metadata är viktiga ur ett användarperspektiv?

I denna studie bedöms datakvalitet främst genom analyser av partiellt bortfall, samt vissa kompletterande indikatorer. Studien är också avgränsad till de data som samlas in och är tillgängliga i registret idag. Undersökning av rapportbortfall och andra frågeställningar som kräver jämförelse med alternativa datakällor ingår inte i denna studie. Undantag till detta är Stradas data om dödade som i denna studie jämförs med officiell statistik från Trafikanalys.

Studien innehåller både kvantitativa och kvalitativa delar och baseras på tre huvudsakliga källor: (1) analys av Stradadata, (2) studie av dokument, och (3) inhämtning av expertinformation. Resultaten visar att täckningen gällande dödade inte är fullgod, användbarheten av kopplade NVDB-data är låg, och att det finns ett behov av bättre dokumentation. Sett till partiellt bortfall indikerar resultaten att kvaliteten ökat för polisrapporter och minskat för sjukvårdsrapporter över tid. Vidare har en beskrivning av matchningsalgoritmen tagits fram, men fullständig korrekthet kan inte garanteras givet bristande dokumentation.

Utifrån beskrivningen och resultaten i denna rapport rekommenderas Transportstyrelsen att:

1. Föra dialog med departement och de berörda myndigheterna Polisen, Trafikverket, Trafikanalys och VTI om konsekvenserna av polisens nedprioritering av rapportering till Strada.
2. Fortsätta att prioritera inkludering av information om personer med skyddade personuppgifter i Strada utan att begränsa tillgången för användarna.
3. Utöka värdemängderna för Stradas variabler för att möjliggöra tydligare indelning av saknade värden.
4. Säkerställa att Stradadata fortfarande kan kopplas till polisens interna utredningsmaterial trots minskningen av angivna diarienummer i Strada.

5. Kartlägga och dokumentera matchningsalgoritmen i detalj för att möjliggöra utvärdering och identifiering av förbättringar.
6. Utveckla en vägledning till användaren som beskriver hur olyckor har kopplats till NVDB och hur kopplingen kan användas.
7. Återinföra vägtyp som polisrapporterad variabel med samma värdemängd som i NVDB.
8. Undersöka möjligheten att utveckla polisens och sjukvårdens rapporteringsverktyg så att endast väglänkar relevanta för olyckan kopplas från NVDB, samt att strukturera dessa data i separata variabler.
9. Dela upp nuvarande användarhandledning för Strada uttagswebb i två dokument där det första innehåller övergripande information, registrets syfte, tidslinje och definitioner, och det andra innehåller information om själva uttagswebben.
10. Skapa en maskinläsbar variabelmetadatabil innehållande namn, typer, värdemängder, startdatum, slutdatum, och förklaringar. Denna bör bifogas till varje datauttag.
11. Erbjud data i textformat som alternativ till Excelformat.

Denna rapport ska inte ses som en fullständig genomlysning av Stradas datakvalitet. Studien studerar inte innehållet i varje variabel i detalj, samt adresserar inte alla datamängder (till exempel data kopplade från Vägtrafikregistret). Datakvalitet är ett vitt begrepp och denna studie fokuserar på vissa specifika aspekter med målet att öka användarbarheten av data från Strada.

Summary

Strada (Swedish Traffic Accident Data Acquisition) is a national register for road traffic crashes that combines data from both the police and all emergency hospitals in Sweden that provide acute care in orthopaedics or surgery. By integrating these data sources, the aim is to provide detailed information on both crash circumstances and injuries for research, evaluation, and target monitoring. Data from Strada constitute a cornerstone for understanding the development of road safety in Sweden and are important for assessing the effects of road safety measures. In addition to police and healthcare reports, the register also includes linked data from the National Road Database (NVDB) and the Road Traffic Register. It is important to examine Strada's data and documentation from a user perspective to ensure that correct trade-offs can be made when using the data. This study has the user perspective as its basis, with the purpose of describing these aspects of Strada's data quality.

The purpose of this study is to describe Strada's data quality from a user perspective and to develop recommendations for improving Strada's data quality and usability. This is done by exploring the following research questions:

- What is the quality of the data reported by the police, and how has this changed over time?
- What is the quality of the data reported by the hospitals, and how has this changed over time?
- How does the matching of police and healthcare reports work?
- How does the linkage to NVDB work, and what information is added to Strada through this connection?
- What documentation and metadata are important from a user perspective?

In this study, data quality is primarily assessed through analyses of partial missing data, along with some complementary indicators. The study is also limited to the data that are currently collected and available in the register. Investigation of underreporting and other issues that require comparison with alternative data sources are not included in this study. An exception is Strada's data on fatalities, which in this study are compared with official statistics from Traffic Analysis.

The methodological approach of the study includes both quantitative and qualitative components and is based on three main sources: (1) analysis of Strada data, (2) document review, and (3) collection of expert information. The results show that the coverage of fatalities is not complete, the usability of linked NVDB data is low, and that there is a need for improved documentation. With regard to partial missing data, the results indicate that the quality of police reports has improved over time, while the quality of healthcare reports has declined. Furthermore, a description of the matching algorithm has been developed, but complete correctness cannot be guaranteed due to insufficient documentation.

Based on the descriptions and results in this report, it is recommended that the Swedish Transport Agency should:

1. Hold discussions with the ministry and the relevant authorities – the Police, the Swedish Transport Administration, Traffic Analysis, and VTI – regarding the consequences of the police's deprioritisation of reporting to Strada.
2. Continue to prioritise the inclusion of individuals with protected personal data in Strada without restricting access for users.
3. Expand the value sets for Strada's variables to enable clearer categorisation of missing values.
4. Ensure that Strada data can still be linked to the police's internal investigation material despite the exclusion of police case reference numbers in Strada.

5. Map and document the matching algorithm in detail to enable evaluation and identification of improvements.
6. Develop a user guide describing how crashes have been linked to NVDB and how this linkage can be used.
7. Reintroduce road type as a police-reported variable with the same value set as in NVDB.
8. Investigate the possibility of further developing the police and healthcare reporting tools so that only road links relevant to the crash are linked from NVDB, and structure these data into separate variables.
9. Split the current user manual for the Strada extraction web service into two documents, where the first contains general information, the purpose of the register, timelines, and definitions, and the second contains information about the extraction web service itself.
10. Create a machine-readable variable metadata file containing names, types, value sets, start dates, end dates, and explanations. This should be included with each data extraction.
11. Offer data in text format as an alternative to the Excel format.

This report should not be regarded as a complete review of Strada's data quality. The study does not include comparisons with other registers, does not examine the content of each variable in detail, and does not address all data sets (for example, data linked from the Road Traffic Register). Data quality is a broad concept, and this study focuses on certain specific aspects with the aim of increasing the usability of the data from Strada.

Förord

Det som räknas räknas. Systematisk insamling av data om de olyckor och skador som sker i vårt vägtransportsystem är grunden till korrekta prioriteringar och utveckling av effektiva trafiksäkerhetsåtgärder. Denna slutrapport är resultatet av ett VTI-lett utvecklingsprojekt i samarbete med Transportstyrelsen som fokuserat på datakvaliteten i den svenska nationella olycksdatabasen Strada.

Jag vill tacka Khabat Amin med kollegor på Transportstyrelsen för ett gott samarbete och för tillhandahållandet av information om Strada och synpunkter på slutrapporten. Tack också till granskare Jenny Eriksson vars kommentarer lett till förbättringar av rapporten.

Linköping, maj 2026

Christian Howard
Projektledare

Granskare/Examiner

Jenny Eriksson, VTI.

De slutsatser och rekommendationer som uttrycks är författarens/författarnas egna och speglar inte nödvändigtvis myndigheten VTI:s uppfattning./The conclusions and recommendations in the report are those of the author(s) and do not necessarily reflect the views of VTI as a government agency.

Slutrapporten är framtagen med ekonomiskt stöd från Skyltfonden, Trafikverket. Ståndpunkter, slutsatser och arbetsmetoder i rapporten reflekterar författaren och överensstämmer inte med nödvändighet med Trafikverkets ståndpunkter, slutsatser och arbetsmetoder inom rapportens ämnesområde.

Publikationen godkänd för publicering/ Publication approved for publication

Ellen Grumert, VTI.

1. Introduktion

Strada (Swedish traffic accident data acquisition) är ett nationellt register för vägtrafikolyckor som kombinerar data från både polisen och samtliga akutsjukhus som ger akutsjukvård inom ortopedi eller kirurgi i Sverige¹. Genom att kombinera dessa datakällor är målet att tillhandahålla detaljerad information om både olycksomständigheter och skador för forskning, utvärdering och måluppföljning. Data från Strada är en grundpelare för förståelsen av trafiksäkerhetsutvecklingen i Sverige och är viktiga för att bedöma effekten av trafiksäkerhetsåtgärder. Utöver polis- och sjukvårdsrapporter inkluderar registret också kopplade data från Nationella vägdatabasen (NVDB) och Vägtrafikregistret.

Datainsamling till Strada påbörjades 1999 och nationell täckning uppnåddes från polisen 2003 och från akutsjukvården 2016. Med Stradas centrala roll i svenskt trafiksäkerhetsarbete är hög datakvalitet av stor vikt. Att upprätthålla hög datakvalitet är också viktigt för att behålla förtroendet för registret. Transportstyrelsen har tidigare publicerat rapporter gällande vissa aspekter av Stradas datakvalitet såsom täckning och bortfall (Fredlund, 2015, Yamazaki, 2018). I dessa konstaterades att Strada inte är heltäckande och att det råder stor variation i rapporteringsgrad mellan trafikantslag och mellan rapporteringskällor. Strada polis och Strada sjukvård har också tidigare jämförts med data från Socialstyrelsens Patientregister (PAR) i Bengtsson (2012) där det uppskattades att Strada fångar upp ungefär 63 procent av alla trafikskadade (givet antagande att Strada och PAR tillsammans fångar upp alla trafikskadade). I Forsman och Eriksson (2023) presenteras en metod (baserad på tidigare arbete utfört av Transportstyrelsen) för att skatta det totala antalet skadade i vägtrafikolyckor, med syfte att justera för varierande täckningsgrad över tid. Genom samarbete mellan Trafikverket, Transportstyrelsen och VTI har denna metod nu arbetats in i Trafikverkets årliga måluppföljning, inte direkt i etappmålsuppföljningen av allvarligt skadade, utan som ett index för att följa utvecklingen över tid (Hurtig, 2024).

Nyligen publicerade Transportstyrelsen en rapport (Fredlund och Frank, 2025) som syftar till att utvärdera vissa effekter av Lag (2021:319) om Transportstyrelsens olycksdatabas. Lagen som infördes 1 juli 2021 innebar flera förändringar för Strada. Bland annat blev rapportering lagstadgad för vårdgivare som ger akutsjukvård inom ortopedi eller kirurgi och därmed upphörde kravet på medgivande från den vårdsökande (vilket innebär att den vårdsökande ska rapporteras även om denne inte fyllt i en så kallad trafikskadejournal). Det innebär också att sjukvårdens rapportering av dödsfall, som upphörde under en period som följd av dataskyddsförordningen (GDPR, EU 2016/679), kunde återupptas. Lagen innebär också att vissa företag och organisationer förlorade sin direktåtkomst till Strada via Strada uttagswebb. I rapporten konstateras att det totala antalet rapporter ökade med 20 procent efter lagens införande. Vidare har tidigare könsskillnader jämnats ut. Samtidigt visar rapporten att endast 42 av de granskade 69 sjukhusen lämnat in rapporter utan trafikskadejournal, vilket tyder på att de övriga 27 inte uppfyller lagkravet. Registreringstiden och antalet rapporter med bristfälliga uppgifter om olycksplatsen har också ökat markant.

Transportstyrelsen har också studerat polisens rapportering. Ett opublicerat PM (Transportstyrelsen, 2025) visar att polisens rapportering till Strada minskat under flera år till följd av att polisen nedprioriterar att åka ut på larm med trafikolycka, att personer med skyddade personuppgifter inte registreras i Strada och att vissa olyckor inte rapporteras till Strada trots att polisen varit på olycksplatsen. Till exempel saknas mellan 1266 och 2381 polisrapporter årligen under åren 2016–2023, trots att polisen varit på plats enligt sjukvården och har ett lagkrav att rapportera samtliga vägtrafikolyckor som kommer till deras kännedom.

¹ Utöver vägtrafikolyckor innehåller Strada fallolyckor i vägtrafikmiljö (vilket rapporteras in från akutmottagningarna). Strada innehåller också olyckor med snöskoter eller terränghjuling utanför vägtrafikmiljö, samt olyckshändelser med fritidsbåtar (dessa rapporteras även av kustbevakningen) men dessa kategorier har dock inte ingått i denna rapport.

Rapportbortfall är dock endast en aspekt av ett registers kvalitet. Partiellt variabelbortfall inom rapporter har berörts i tidigare studier och rapporter, men systematiska genomgångar är inte tillgängliga. Det är också till viss del oklart hur matchningen mellan polis- och sjukvårdsrapporter fungerar. Detta för att det saknas både dokumentation om hur matchningen går till och utvärderingar av hur väl den fungerar (metodikerna för matchningen har inte utvärderats sedan Björketun, 2006). Genomgångar av kopplingen av olyckor till NVDB har heller inte publicerats tidigare.

Det är viktigt att granska Stradas data och dokumentation ur ett användarperspektiv, så att välgrundade avvägningar kan göras kring eventuella brister och biaser i data. Denna studie tar sin utgångspunkt i detta användarperspektiv med syfte att beskriva dessa aspekter av Stradas datakvalitet.

2. Syfte

Syftet med denna studie är att beskriva Stradas datakvalitet ur ett användarperspektiv och att ta fram rekommendationer för att förbättra Stradas datakvalitet och användarbarhet. Detta genom att utforska följande frågeställningar:

- Vilken kvalitet har data som rapporteras in från polisen och hur har detta förändrats över tid?
- Vilken kvalitet har data som rapporteras in från sjukvården och hur har detta förändrats över tid?
- Hur fungerar matchningen av polis- och sjukvårdsrapporter?
- Hur fungerar kopplingen mot NVDB och vilken information tillförs Strada genom denna koppling?
- Vilken dokumentation och vilka metadata är viktiga ur ett användarperspektiv?

I denna studie bedöms datakvalitet främst genom analyser av partiellt bortfall, samt vissa kompletterande indikatorer. Studien är också avgränsad till de data som samlas in och är tillgängliga i registret idag. Undersökning av rapportbortfall och andra frågeställningar som kräver jämförelse med alternativa datakällor ingår inte i denna studie. Undantag till detta är Stradas data om dödade som i denna studie jämförs med officiell statistik från Trafikanalys.

3. Metod

Studiens metodansats innehåller både kvantitativa och kvalitativa delar och baseras på tre huvudsakliga källor: (1) analys av Stradadata, (2) studie av dokument och (3) inhämtning av expertinformation.

3.1. Stradadata

Analyserna av polis-, sjukvårds- och NVDB-data baseras på ett datauttag för åren 2016–2024. Data hämtades från Strada uttagswebb 17 oktober 2025. År 2016 valdes som startår eftersom detta var första året då samtliga vårdgivare som ger akutsjukvård inom ortopedi eller kirurgi var anslutna till Strada. Data bestod av 397 738 olyckor och 522 497 personer. För jämförelse hämtades också officiell statistik om antalet dödade från Trafikanalys hemsida² för samma tidsperiod.

Analyserna av matchningen av rapporter baseras på datauttag för åren 2022 och 2023. Data hämtades från Strada uttagswebb 7 september 2023 respektive 23 november 2024. Dessa år valdes eftersom en ny lag som påverkade rapporteringsgraden hos sjukvården infördes 2021. Olyckstyperna G0 (fotgängare singel) och G8 (fotgängare mot fotgängare) exkluderades då de endast rapporteras från sjukvården. Även dödsolyckor exkluderades. Data bestod då av 67 971 olyckor och 93 427 personer. Utöver att beskriva hur matchningen går till undersöks likheter och skillnader mellan matchade rapporter. Matchningsgrad studeras inte i den här rapporten men beskrivs närmare i en artikel som skickats in till en vetenskaplig tidskrift (Forsman och Howard, 2026).

Analysmetodiken bestod av statistiska beräkningar och kontroller, samt deskriptiva statistiska sammanställningar. Behovet av dokumentation och metadata kartlades parallellt med analysen utifrån frågor och oklarheter som uppstod under arbetets gång.

3.2. Dokument

Följande dokument har erhållits från Transportstyrelsen:

- Dokument från AerotechTelub. STRADA, SRS Matchning. ADA132A. Datum 29 augusti 2003. Detta dokument innehåller ett flödesschema över matchningen och vissa beräkningsformler. Internt dokument.
- Dokument från Triona. Sammanfattning av matchningsfunktionen. Datum 3 september 2015. Detta dokument innehåller en delvis beskrivning av matchningen och viss SQL-kod. Internt dokument.
- Variabellista. Datum 8 september 2023. Exceldokument som används vid beställning av Stradadata. Externt dokument tillgängligt på Transportstyrelsens hemsida³.
- Strada poliswebb –Handledning augusti 2025. Datum augusti 2025. Användarhandledning för polisens webbrapportering. Externt dokument tillgängligt på Transportstyrelsens hemsida.
- Strada uttagswebb –Handledning till Strada uttagswebb version 4.0. Datum ej angivet. Användarhandledning för Strada uttagswebb innehållande viss information om variabler och datastruktur. Externt dokument tillgängligt på Transportstyrelsens hemsida³.

Då dessa dokument är knapphändiga i vissa delar och även daterade i vissa fall har de endast använts som en bas för förståelse. Kontroller har därefter utförts i Strada i de fall det varit möjligt.

² <https://www.trafa.se/vagtrafik/vagtrafikskador/>

³ <https://www.transportstyrelsen.se/STRADA>

3.3. Expertinformation

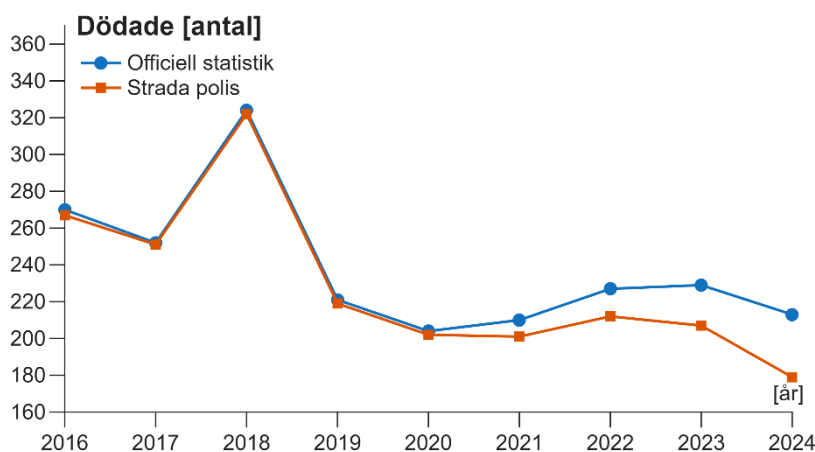
Expertinformation har inhämtats genom kommunikation med Khabat Amin, statistiker och trafiksäkerhetsanalytiker på avdelning Väg- och järnväg, på Transportstyrelsen, och Transportstyrelsens IT-förvaltning. Kommunikation har skett både muntligt och skriftligt. Processen har varit iterativ där förfrågningar om information och återkoppling av resultat skett i omgångar medan projektet pågår.

4. Resultat

Resultatavsnittet är uppdelat i sex underavsnitt. Först redovisas resultat om rapporteringen av dödade i Strada (4.1). Därefter behandlas polisrapporterade data (4.2), sjukvårdsrapporterade data (4.3), rapportmatchning (4.4), och NVDB-kopplade data (4.5) i varsitt avsnitt. Avslutningsvis presenteras resultat kring Stradas dokumentation och metadata (4.6).

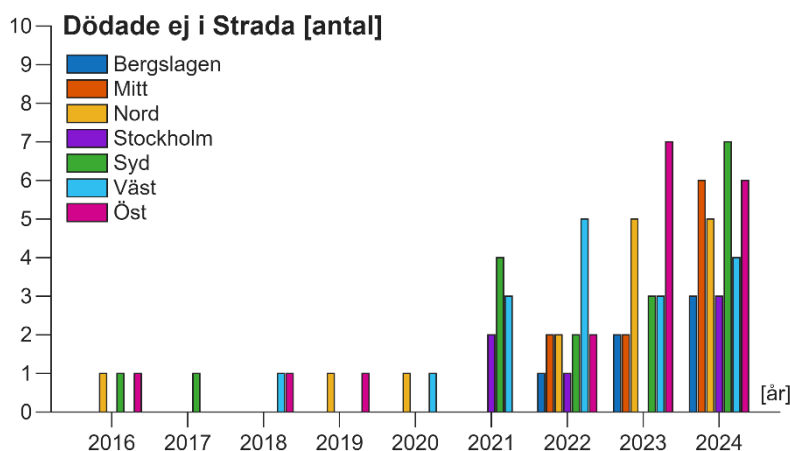
4.1. Dödade enligt polis och sjukvård

För antalet dödade kan Stradadata jämföras med officiell statistik från Trafikanalys. Polisen har ansvaret för att rapportera samtliga dödsfall som kommer till deras kännedom, men Figur 1 visar att ett antal dödsfall inte är registrerade i Strada. För perioden 2016–2020 saknas enstaka dödsfall, men från och med 2021 ökar antalet saknade fall successivt. År 2024 var skillnaden den största hittills där 34 dödade personer saknas i Strada. Enligt Transportstyrelsen har de saknade fallen främst tre orsaker: (1) att personer med skyddad identitet inte registreras i Strada, (2) minskad polisrapportering av personer som dör i efterhand på sjukhus inom 30 dagar och (3) att polisen av oklar anledning ibland inte rapporterar till Strada trots att de har upprättat en trafikmålsanteckning.



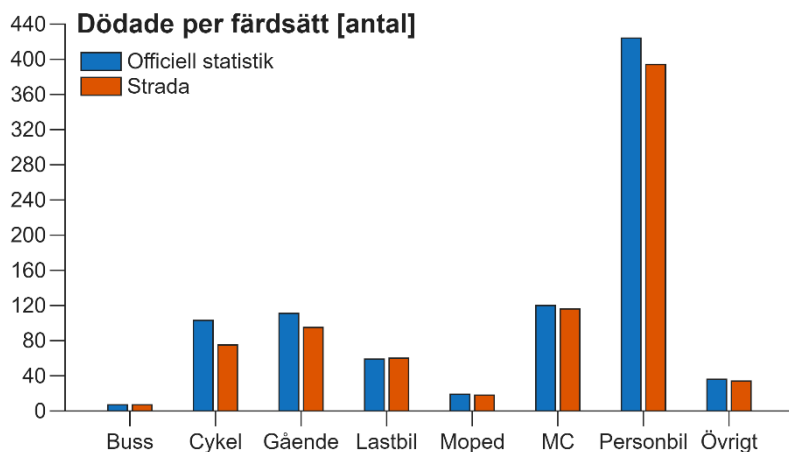
Figur 1. Antalet dödade enligt officiell statistik från Trafikanalys jämfört med Strada.

I Figur 2 redovisas antalet personer som saknas i Strada uppdelat på polisregion och år. Figuren visar att alla polisregioner har bortfall under perioden 2021–2024. Det är däremot svårt att utläsa några tydliga skillnader mellan polisregioner då faktiska antalet dödade kan variera mycket från år till år.



Figur 2. Antalet dödade som saknas i Strada uppdelat på polisregion för 2021–2024.

Figur 3 visar totalt antal dödade enligt officiell statistik jämfört med Strada per färdssätt för 2021–2024. Figuren visar att det saknas flest bilister (30 personer), cyklister (28 personer) och gående (16 personer) i Strada jämfört med officiell statistik. Cyklister och gående verkar också sakna oproportionerligt många fall jämfört med bilister: 27,2 procent saknade cyklister och 14,4 procent saknade gående jämfört med 7,1 procent saknade bilister. För lastbil innehåller Strada ett fall mer än den officiella statistiken. Vad detta beror på är oklart.

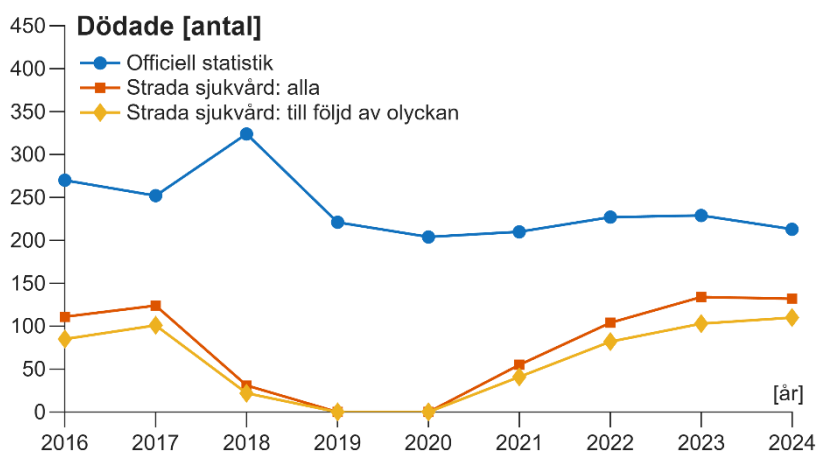


Figur 3. Totalt antal dödade enligt officiell statistik jämfört med Strada per färdssätt för 2021–2024.

För aktuell period har sjukvården registrerat dödade som kommer in via akuten enligt följande underkategorier:

- Död vid ankomst, annan orsak
- Död vid ankomst, osäker orsak
- Död vid ankomst, till följd av olyckan
- Död, annan orsak
- Död, osäkert
- Död, till följd av olyckan.

I Figur 4 jämförs det officiella antalet dödade personer med antalet registrerade av sjukvården, både totalen för alla sex underkategorier och totalen för de två underkategorier där sjukvården angett att dödsfallet var till följd av olyckan. Figuren visar att sjukvården successivt återupptagit rapporteringen av dödade efter uppehållet 2019–2020 och att sjukvården under 2024 fångat ungefär hälften så många dödade jämfört med den officiella statistiken.



Figur 4. Antalet dödade personer registrerade i Strada sjukvård jämfört med officiell statistik från Trafikanalys.

I Tabell 1 redovisas fördelningen av sjukvårdens rapporterade skadegrad för de som rapporterats som död enligt officiell statistik av polisen. Det framgår att en hög andel, 76,8 procent, av polisrapporterade dödade inte rapporterats av sjukvården. Det framgår också att det finns vissa fall där polisen och sjukvården inte är samstämmiga om dödsfallet ska räknas som en vägtrafikolycka - fall där sjukvården inte angett att dödsfallet var till följd av olyckan eller har angett en annan skadegrad än död. Med fokus på fallen där sjukvården rapporterat död till följd av olyckan visar Tabell 2 fördelningen av polisens rapporterade skadegrad. Det framgår att en hög andel, 72,8 procent, också rapporterats som dödade från polisen. Samtidigt är 19,3 procent inte rapporterade av polisen och det finns också fall där klassificeringen är suicid eller svårt skadad.

Sammantaget visar resultaten angående antalet dödade att Strada inte är heltäckande gällande dödsfall i vägtrafiken som följd av polisens minskade rapportering. Sjukvården kan heller inte kompensera för detta bortfall eftersom denna rapportering inte varit stabil, annan värdemängd används, och oklarheter råder när i processen sjukvårdens rapportering sker.

Tabell 1. Fördelning av sjukvårdsrapporterad skadegrad för personer som rapporterats som död enligt officiell statistik av polisen för 2016–2024.

Skadegrad sjukvård	Antal	Andel
<i>Ej rapporterad av sjukvården</i>	1 583	76,8 %
Död vid ankomst, till följd av olyckan	201	9,8 %
Död, till följd av olyckan	195	9,5 %
Allvarligt skadad (ISS 9-)	27	1,3 %
Död, osäkert	21	1,0 %
Död vid ankomst, osäker orsak	19	0,9 %
Måttligt skadad (ISS 4-8)	6	0,3 %
Död, annan orsak	4	0,2 %
Okänd skada	3	0,1 %
Lindrigt skadad (ISS 1-3)	1	0,0 %
Total	2 060	100,0 %

Tabell 2. Fördelning av polisrapporterad skadegrad för personer som rapporterats som "Död vid ankomst, till följd av olyckan" eller "Död, till följd av olyckan" av sjukvården för 2016–2024.

Skadegrad polis	Antal	Andel
Död	396	72,8 %
Ej rapporterad av polisen	105	19,3 %
Död, ej officiell statistik	36	6,6 %
Svårt skadad	7	1,3 %
Total	544	100,0 %

4.2. Polisrapporterade data

Polisrapporterade data delades in i två kategorier: olycksvariabler och personvariabler. Först undersöktes hur andelen saknade data per variabel utvecklats över tid. Att notera för denna analys är att det saknas både systematik och dokumentation om hur saknade data anges i datauttagen från Strada uttagswebb. De fyra relevanta variabelvärden som identifierades är "Okänt", "Okänd", "Uppgift saknas", eller att det bokstavligen saknas ett angivet värde (alltså att cellen i Excelarket är tom). Det sistnämnda refererar vi till som "Tom" härnäst i denna rapport. Hur dessa värden ska tolkas är dock inte självklart. Vissa variabler har till exempel bara två olika värden: "Tom" eller "Ja". I dessa fall går det inte att utläsa om "Tom" betyder: (1) polisen har beaktat variabeln i sin rapportering och svaret är nej), (2) polisen har beaktat variabeln och svaret är okänt, (3) polisen har missat att rapportera värdet för denna variabel, eller (4) variabeln är inte relevant för aktuell olycka. Samma problematik uppstår för variabler som kan ha värdet "Uppgift saknas" eller "Okänt", men också värdet "Tom".

Tabell B1.1 och Tabell B1.2 i Bilaga 1 anger andelen "Okänt", "Okänd", "Uppgift saknas", och "Tom" per olycksvariabel respektive personvariabel och år. För vissa specifika personvariabler inkluderas även "Okänt fordon", "Övrig/okänd plats" och "Okänd skadegrad" som saknade data. Tabellerna visar att vissa variabler, till exempel *År*, *Månad*, *Dag*, *N-sweref*, *E-sweref*, *Län*, *Kommun*, *Händelseförlopp*, *Olyckstyp* och *Olycksplats*, alltid har ett angivet värde (0 procent saknade värden). Andra variabler innehåller aldrig, eller väldigt sällan med hänsyn till avrundning, någon information (100 procent saknade värden). Det är svårt att utläsa från data för vilka variabler en hög andel saknade värden är en indikation på ett problem med datakvaliteten. Till exempel är nära 100 procent saknade värden för *Osäkert datum* troligtvis en följd av att datum i princip alltid är angivet, medan 99–100 procent för *Stod stilla/parkerat*, *U-sväng* eller *Övningskörning* kan betyda att dessa företeelser är ovanliga eller att variablerna inte används av polisens.

Noterbart i Tabell B1.1 och Tabell B1.2 är att andelen saknade värden minskat årligen för *Hastighetsbegränsning väg A*, från 28 procent år 2016 till 13 procent år 2024. Andelen har även minskat för andra variabler som rör vägens omständigheter, såsom *Hastighetsbegränsning väg B*, *Ljusförhållande*, *Trafikanvisning väg A/B*, *Trafikanvisning väg A/B*, *Trafikreglering A/B*, *Trafiksignal A/B*, *Väderlek*, och *Väglag*. Variablerna *Vägtyp A/B* går dock i motsatt riktning där värden saknas helt från 2021 och framåt. Detta är en följd av ändringar i polisens rapporteringsrutiner och behandlas separat i avsnitt 4.5.

Andelen rapporter som saknar diarienummer har ökat markant från och med 2020, där inga rapporter saknade diarienummer före 2020 och 56 procent saknade diarienummer 2024. Enligt Transportstyrelsens beror detta på att de ersatts med ett TRAFONummer när polisen rapporterar via TRAFON (en app som polisen använder vid rapportering till Strada). Ett annat tydligt skifte ses hos *Fyrhjuling* där andelen saknade värden är 100 procent före 2020, men minskar sedan snabbt till 0 procent från och med 2022. I det här fallet beror detta på att *Fyrhjuling* införts som ny variabel.

4.3. Sjukvårdsrapporterade data

Sjukvårdsrapporterade data analyserades avseende saknade data på samma sätt som för polisrapporterade data. I det här fallet identifierades följande relevanta variabelvärden: ”Tom”, ”Okänt”, ”Okänd”, ”Uppgift saknas”, ”Övrig/okänd plats”, ”Okänd skada” och ”Värde mängd saknas”. Analogt med polisdata saknas information om hur ”Tom” (tom Excelcell i datauttaget) ska tolkas.

Tabell B2.1 i Bilaga 2 anger andelen saknade data per variabel och år. Tabellen visar att vissa variabler; *År, Månad, Dag, Inkom med, Kön, Olyckstyp, Rapport-ID, Registreringstid (antal dagar), Sjukhus, Trafikantkategori – Huvudgrupp, Trafikantkategori – Undergrupp, Vårdförlopp*, och *ålder*, alltid har ett angivet värde (0 procent saknade värden). Andra variabler, specifikt *Avliden datum, Okänt datum, Dödsorsak, Osäkert datum*, innehåller aldrig, eller väldigt sällan med hänsyn till avrundning, någon information (100 procent saknade värden). På samma sätt som med vissa polisvariabler kan det vara svårtolkat om en hög andel saknade värden är en indikation på ett problem med datakvaliteten. Till exempel, kan 98–99 procent för *Barnstol/kudde* innebära att förekomsten är ovanlig.

För flera variabler ökar andelen saknade värden från och med 2020. Detta gäller *Bebyggelsestyp, Bälte, E-sweref, Hjälms, I Konflikt med – Huvudgrupp, I Konflikt med – Undergrupp, ISS, Klockslag, Kommun, Krockkudde utlöst, Län, MAIS, N-sweref, Nya ISS, Olyckan inträffade, Olycksplats, Olycksväg/-gata, Platstyp, Trafikantroll*, och *Vägomständigheter*. Många av dessa variabler rör olyckans plats och omständigheter och detta indikerar därför en minskad kvalitet i detta avseende. Noterbart är att olyckans GNSS-position (*E-sweref* och *N-sweref*), *Vägomständigheter, Olyckan inträffade* (som rör arbete/fritid) ligger på nivåer kring 20 procent saknade värden efter 2020 jämfört med cirka 5 procent före 2020. Även för *Klockslag* har andelen saknade värden ökat märkbart, från cirka 4 procent till cirka 15 procent.

Skademåtten MAIS och ISS som baseras på sjukvårdens AIS-klassificering av skador har gått från 0 procent saknade värden före 2020 till cirka 10 procent saknade värden efter 2020. Detta verkar dock bero på förändringar av hur personer som är oskadade, har okänd skada, eller har osäker skada kodas i systemet. Tabell 3 och Tabell 4 visar antalet personer med MAIS-värde noll respektive ”Tom” (saknas värde) per skadegrad och år. Jämförelse mellan tabellerna visar ett skifte 2020 från att oskadade, okända, och osäkra skador får MAIS-värde noll till att de får värdet ”Tom”. Tabellerna visar också att vissa personer med skadegrad död kodats med värde noll eller ”Tom”. Andelen personer med *Skadegrad* ”Okänd skada”, ”Oskadad” eller ”Osäker skada” ligger relativt stabilt över åren, vilket visas i Tabell 5.

Tabell 3. Antalet personer med MAIS-värde noll per skadegrad och år.

Skadegrad	2016	2017	2018	2019	2020	2021	2022	2023	2024
Död vid ankomst, annan orsak	1	2	1	0	0	0	1	1	1
Död vid ankomst, osäker orsak	4	2	2	0	0	3	4	2	1
Död vid ankomst, till följd av olyckan	21	25	11	0	0	5	15	26	18
Död, annan orsak	1	1	1	0	0	0	0	0	0
Död, till följd av olyckan	0	1	0	0	0	0	0	0	0
Okänd skada	333	271	281	281	85	2	3	2	4
Oskadad	3 990	4 043	3 733	4 066	1 424	0	0	0	0
Osäker skada	229	149	192	182	79	0	0	0	0

Tabell 4. Antalet personer med MAIS-värde "Tom" per skadegrad och år.

Skadegrad	2016	2017	2018	2019	2020	2021	2022	2023	2024
Död vid ankomst, osäker orsak	0	0	0	0	0	0	0	1	0
Död vid ankomst, till följd av olyckan	0	0	0	0	0	2	2	4	5
Död, annan orsak	0	0	0	0	0	0	0	2	0
Död, till följd av olyckan	0	0	0	0	0	0	0	3	0
Okänd skada	0	0	0	1	53	306	423	408	503
Oskadad	1	13	7	8	868	3 025	4 158	4 133	3 452
Osäker skada	0	7	0	0	46	218	236	203	165

Tabell 5. Andel Okänd skada, Oskadad eller Osäker skada per sjukvårdsrapport för variabeln Skadegrad.

Skadegrad	2016	2017	2018	2019	2020	2021	2022	2023	2024
Okänd skada	1 %	1 %	1 %	1 %	1 %	1 %	1 %	1 %	2 %
Oskadad	10 %	10 %	10 %	12 %	9 %	9 %	10 %	9 %	7 %
Osäker skada	1 %	0 %	1 %	1 %	0 %	1 %	1 %	0 %	0 %

För variabelerna som rör onykterhet är andelen saknade värden inte stabil över tid. För *Alkoholpåverkan* varierar andelen mellan 47 procent och 90 procent, medan för *Drogpåverkan* varierar andelen mellan 51 procent och 100 procent, beroende på år. För att ytterligare belysa svarsfördelningen i dessa variabler visas antal sjukvårdsrapporter per svarsalternativ för *Alkoholpåverkan* i Tabell 6 och för *Drogpåverkan* i Tabell 7. Utifrån dessa tabeller ser det ut som att variabeln *Drogpåverkan* införs år 2020 och att detsamma gäller svarsalternativet Misstanke för variabeln *Alkoholpåverkan*. Dock förekommer enstaka "Nej"-svar för variabeln *Drogpåverkan* redan före 2020. Det är också oklart i hur stor utsträckning variabelerna faktiskt används även efter 2020 eftersom den oftast lämnas tom. År 2021 ändras svarmönstret då antal "Nej"-svar är betydligt högre än både före och efter det året.

Tabell 6. Antal sjukvårdsrapporter per svarsalternativ för variabeln alkoholpåverkan.

Trafikantkategori	Alkohol-förekomst	2016	2017	2018	2019	2020	2021	2022	2023	2024
Personbilsförare	Ja, provsvar	103	90	64	36	56	80	118	86	80
Personbilsförare	Misstanke	0	0	0	0	5	18	31	35	17
Personbilsförare	Nej	2 048	1 879	1 596	1 259	1 316	3 203	1 538	649	626
Personbilsförare	Tom	6 978	6 638	5 597	5 234	3 010	2 373	4 979	6 253	5 750
Cyklist	Ja, provsvar	205	199	176	152	153	324	568	482	573
Cyklist	Misstanke	0	0	0	1	99	497	783	710	822
Cyklist	Nej	1 843	1 834	1 777	1 779	2 556	5 968	1 990	597	518
Cyklist	Tom	8 219	8 229	7 671	7 742	6 106	4 909	11 956	12 992	14 140

Tabell 7. Antal sjukvårdsrapporter per svarsalternativ för variabeln drogpåverkan.

Trafikant-kategori	Drog-förekomst	2016	2017	2018	2019	2020	2021	2022	2023	2024
Personbilsförare	Ja, provsvar	0	0	0	0	3	22	42	34	34
Personbilsförare	Misstanke	0	0	0	0	6	18	41	29	32
Personbilsförare	Nej	2	1	5	5	813	3 012	1 239	363	325
Personbilsförare	Tom	9 127	8 606	7 252	6 524	3 565	2 622	5 344	6 597	6 082
Cyklist	Ja, provsvar	0	0	0	0	2	11	18	23	26
Cyklist	Misstanke	0	0	0	0	9	30	55	34	47
Cyklist	Nej	2	7	2	4	1 683	6 257	1 924	515	446
Cyklist	Tom	10 265	10 255	9 622	9 670	7 220	5 400	13 300	14 209	15 534

4.4. Matchning av rapporter

Polisen rapporterar på olycksnivå vilket innebär att de skickar in en rapport som innehåller både övergripande information om olyckan och information som är specifik för varje person. De personer som rapporteras är förare av motorfordon, cyklister och fotgängare. Övriga inblandade i olyckor rapporteras bara om polisen bedömer att de är skadade. Sjukvården rapporterar endast på personnivå, vilket innebär separata rapporter för varje person som besöker en akutmottagning. Ingen koppling görs av sjukvården mellan personer som varit med i samma olycka.

I samband med att polis- och sjukvårdsrapporter arkiveras i Strada görs sedan en matchning, dels mellan sjukvårdsrapporter för att koppla ihop personer som varit med i samma olycka, dels mellan polis- och sjukvårdsrapporter för att koppla ihop personer som rapporterats från båda källorna eller olika personer som varit med i samma olycka men är inrapporterade från olika källor. Enligt Transportstyrelsen kan rutinerna för när arkiveringen sker skilja sig mellan olika instanser, till exempel kan det ske löpande eller månadsvis. Matchningen görs automatiskt i programvaran men i vissa fall görs också en manuell kontroll för att avgöra om det ska ske en matchning eller inte.

Matchningen kan sägas vara antingen på personnivå eller olycksnivå. Personnivå innebär att en person som varit inblandad i en olycka har rapporterats från både polis och sjukvård. För att det ska ske en

matchning på personnivå ska matchningsalgoritmen identifiera att det är samma person och samma olyckstillfälle. Olycksnivån kan antingen innebära att två personer från samma olycka har rapporterats in från sjukvården, eller att en eller flera personer från sjukvården kan kopplas ihop med en olycka som rapporterats in från polisen men inte med specifika personer. Det senare kan till exempel uppstå om polisen rapporterar in en olycka där en, enligt polisen, oskadad passagerare är inblandad. Denna person kommer då inte att rapporteras av polisen. Om personen ändå uppsöker en akutmottagning och rapporteras från sjukvården kan den kopplas till olyckan om den uppfyller vissa kriterier. En liknande situation kan uppstå om personen faktisk är rapporterad av både sjukvård och polis men där det inte går att säkerställa att det är samma person på grund av saknad information.

Nedan beskrivs mer i detalj hur matchningen sker på person- och olycksnivå. Beskrivningen är baserad på de källor som nämns i metodkapitlet.

4.4.1. Matchning på personnivå

För att det ska ske matchning på personnivå måste båda rapporterna (från polis och sjukvård) innehålla fullständiga personnummer⁴. Förutom att personnummer ska vara detsamma i de två rapporterna går matchningen till på följande sätt.

1. Om tiden för händelsen skiljer sig med högst 24 timmar så sker en matchning, oavsett hur långt avståndet är mellan de angivna platserna. I de fall där avståndet är längre än 1000 meter och/eller tidsskillnaden är mellan 20–24 timmar görs dock en manuell bedömning för att avgöra om matchningen ska avvisas eller inte.
2. Om det finns flera matchningar enligt 1. matchas det par av rapporter som har högst Q-värde (se formler nedan).
3. Om klockslag saknas sätts värdet till 00:00 samma datum och sedan baseras matchningen på denna tidsangivelse.

Q-värdet beräknas på olika sätt beroende på vilken information som finns tillgänglig. Nedanstående formel⁵ gäller då åtminstone personnummer och tid finns

$$Q = \frac{k_a Q_a + k_t Q_t + k_p Q_p}{k_a + k_t + k_p} \quad (1)$$

där konstanterna, k , har följande värden: $k_a = 1$, $k_t = 3$, $k_p = 8$. $Q_p = 1$ om personnumret är fullständigt, vilket är en förutsättning för matchning på personnivå. Vidare beräknas Q_a som

$$Q_a = \begin{cases} 0 & \text{om } x_{diff} > 1000 \text{ m eller } y_{diff} > 1000 \text{ m} \\ 1 - \sqrt{\frac{x_{diff}^2 + y_{diff}^2}{2 \cdot 1000^2}} & \text{annars} \end{cases} \quad (2)$$

där x_{diff} och y_{diff} är avstånd i x-led respektive y-led mellan rapporternas platsangivelse. Q_a sätts också till 0 om platsangivelse saknas. Slutligen beräknas Q_t som

⁴ Samordningsnummer används också på samma sätt som personnummer för matchningen i Strada. Samordningsnummer tilldelas personer som inte är folkbokförda i Sverige men ändå behöver kunna identifieras av myndigheter.

⁵ Dokumentationen är inte konsekvent, där Q-värdet i Ekvation 1 ligger mellan 0 och 1 men i Ekvation 4 ligger mellan 0 och 100. Detta har tolkats att betyda samma sak.

$$Q_t = \begin{cases} 1 & \text{om } t_{diff} \leq 1,2 \\ 0,8 & \text{om } 1,2 < t_{diff} \leq 2,4 \\ 0,6 & \text{om } 2,4 < t_{diff} \leq 3,6 \\ \frac{1}{85} \left(40 - \frac{t_{diff}}{10T} \right) & \text{om } 3,6 < t_{diff} < 24,0 \end{cases} \quad (3)$$

där $T = 0,75$ och t_{diff} är differensen i tid mellan rapporterna där tiden mäts i timmar, 1,2 timmar motsvarar alltså 72 minuter.

Observera att i Strada anges bara ett Q-värde per olycka trots att Q beräknas för varje par av rapporter. Det Q som anges är från det par av rapporter som har högst Q-värde.

4.4.2. Matchning på olycksnivå, endast sjukvårdsrapporter

Matchning för att avgöra om två sjukhusrapporter kommer från samma olycka baseras på uppgifter om position, tid och olyckstyp. Om olyckstyp är G0 (fotgängare singel) eller G1 (cykel singel) sker ingen matchning och olyckstyp S (singel motorfordon) matchas bara mot andra rapporter med samma olyckstyp. I övrigt sker matchning om Q-värdet är minst 65. I detta fall beräknas Q-värdet baserat på position och tid på följande sätt:

$$Q = \frac{100 + Q_{min} \frac{k_a Q_a + k_t Q_t}{k_a + k_t}}{2} \quad (4)$$

där $Q_{min} = 65$, $k_a = 1$, $k_t = 3$, Q_a beräknas enligt (2) och Q_t enligt (3).

Kravet att Q ska vara minst 65 betyder i praktiken att Q_t behöver vara minst 0,8 och därmed får differensen i tid vara högst 2,4 timmar (kan även krävas kortare tid beroende på avstånd mellan platserna). På samma sätt måste Q_a vara minst 0,152 vilket till exempel fås om differensen är cirka 850 m i både x- och y-led.

4.4.3. Matchning på olycksnivå, polis- och sjukvårdsrapporter

Om fullständigt personnummer saknas sker ingen matchning på personnivå mellan polis- och sjukvårdsrapporter, men däremot kan rapporter kopplas till samma olycka. Dessa kommer då att ha samma olycksnummer (*Olycksnummer*) men inte samma referensnummer (*Referens*). Q-värdet beräknas på samma sätt som vid matchning mellan sjukvårdsrapporter enligt (4) och även här sker matchning om Q är minst 65. Vidare matchas olyckstyperna S, G1 och G2 (moped singel) bara mot samma olyckstyp. G0 matchas inte alls (dessa rapporteras endast från sjukvården).

4.4.4. Resultat av matchning i Strada år 2022 och 2023

Datamaterialet innehåller alla olyckor som rapporterats under 2022 och 2023 förutom dödsolyckor och förutom olyckor av typ G0 och G8. Totalt är det 67 971 olyckor.

För att visa hur ofta matchning sker har olyckorna delats in efter om de innehåller polis- och/eller sjukvårdsrapporter och om matchning skett samt vilken typ av matchning (Tabell 8). Den största gruppen är olyckor med endast en sjukvårdsrapport (som motsvarar en person) som inte matchats med vare sig polisrapport eller annan sjukvårdsrapport. Det är känt sedan tidigare att många olyckor med cykel återfinns i denna grupp (Amin et al. 2022). Detta är en stor olycksgrupp som oftast bara rapporteras från sjukvården, speciellt när det gäller singelolyckor. Den näst största gruppen är olyckor med endast polisrapport, men till skillnad från sjukvårdsrapporterna kan dessa innefatta flera personer eftersom polisen rapporterar på olycksnivå. I de fall där det skett en matchning är det vanligast att kopplingen är på personnivå, det vill säga att en sjukvårdsrapport kopplats ihop med en person från en polisrapport med hjälp av personnummer. I 9 856 olyckor finns det en eller flera personkopplingar

men inga andra kopplingar. I 805 olyckor finns det rapporter från både sjukvård och polis men endast kopplingar på olycksnivå. Det vill säga att de personer som är inblandade i olyckan är rapporterade från antingen polisen eller sjukvården men inte från båda, eller att de är rapporterade från båda källorna men inte kan kopplas samman på grund av saknade eller ofullständiga personnummer. Det senare är förmodligen vanligt då en stor andel av de som redovisas har annan nationalitet än svensk, det gäller för 47,5 procent av de polisrapporterade personerna (nationalitet anges inte i sjukvårdsrapporterna). Sedan förekommer också 1 076 olyckor som endast rapporterats från sjukvården men där två eller flera sjukvårdsrapporter kopplats ihop till samma olycka. Observera att olyckorna som ligger till grund för Tabell 8 innehåller alla skadegrader utom dödsolyckor, även en del olyckor utan personskada. Detta kan påverka om matchning sker eller inte eftersom det är mindre sannolikt att personer i dessa olyckor besöker en akutmottagning.

Tabell 8. Antal olyckor uppdelat efter om och hur matchning mellan rapporter har skett.

Typ av rapporter	Typ av matchning	Antal olyckor
Olyckor med både polis- och sjukvårdsrapporter	Endast personnivå	9 856
Olyckor med både polis- och sjukvårdsrapporter	Endast olycksnivå	805
Olyckor med både polis- och sjukvårdsrapporter	Både person- och olycksnivå	508
Olyckor med endast polisrapporter	Ingen matchning	12 866
Olyckor med endast sjukvårdsrapporter	Olycksnivå	1 076
Olyckor med endast sjukvårdsrapport	Ingen matchning	42 845
Osäker information	-	15
Total		67 971

I Tabell 9 visas antal rapporterade personer från olika källor. De personer som rapporterats från både polis och sjukvård är de som matchats på personnivå. Att antal matchade på personnivå är något högre i Tabell 9 än i Tabell 8 beror på att det kan finnas mer än en personmatchning per olycka.

Tabell 9. Antal rapporterade personer uppdelat efter rapporteringskälla.

Rapporteringskälla	Antal personer
Endast polis	34 436
Endast sjukvård	46 876
Polis och sjukvård (matchning på personnivå)	12 115
Total	93 427

I Tabell 10 visas hur ofta det saknas värden på de variabler som används i matchningen. När det gäller personnummer, som är den viktigaste variabeln för att identifiera om två rapporter avser samma person, finns inte uppgift om personnummer är komplett eller inte i uttag från uttagswebben som vi använt här. Denna variabel finns därför inte med i tabellen. I övrigt kan man konstatera att datum och olyckstyp alltid finns angivet, i både polis- och sjukvårdsrapporter. Däremot saknas klockslag och position relativt ofta från sjukvården. Detta är främst ett problem i de fall där det saknas personnummer eller till exempel där olika personer i samma olycka har rapporterats endast från sjukvården. I dessa fall görs ingen matchning om det saknas värde på position. Om det saknas klockslag sätts detta till 00:00, så i vissa fall kan det ske en matchning, men det beror på när olyckan faktiskt skett.

Tabell 10. Antal och andel saknade värden samt antal osäkra värden för variabler som är med i matchningsalgoritmen. Baserat på rapporterade personer. Totalt antal rapporterade från polis = 46 551 och totalt antal rapporterade från sjukvård = 58 991.

Variabel	Rapporteringskälla	Antal saknade värden	Andel saknade värden	Antal osäkra värden
Datum	Polis	0	0,0 %	4
Datum	Sjukvård	0	0,0 %	157
Klockslag	Polis	276	0,6 %	122
Klockslag	Sjukvård	8 621	14,6 %	3 813
Position (koordinater)	Polis	0	0,0 %	3 374
Position (koordinater)	Sjukvård	12 663	21,5 %	9 841
Olyckstyp	Polis	0	0,0 %	NaN
Olyckstyp	Sjukvård	0	0,0 %	NaN

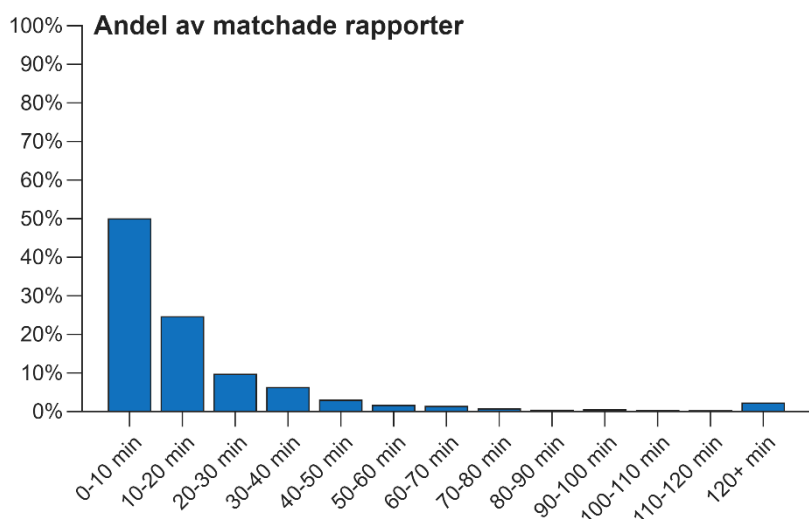
Skillnad i tid, avstånd och olyckstyp i matchade rapporter

I det här avsnittet visas faktiska skillnader i tid, avstånd och olyckstyp mellan rapporter som matchats på personnivå. Alltså där samma person rapporterats från både polis och sjukvård och där matchning skett.

Av totalt 12 115 matchade par saknas klockslag från sjukvård eller polis i 1 214 fall. Dessa har i matchningsalgoritmen ersatts med 00:00 men i Tabell 11, som visar antal och andel matchande par efter tidsavstånd, är de exkluderade. Överensstämmelsen mellan polis och sjukvård är mycket god, i 94,7 procent av paren är skillnaden i tidsangivelse mindre än en timme. När tiden delas in i kortare intervall framkommer också att i hälften av alla par är skillnaden mindre än 10 minuter (Figur 5).

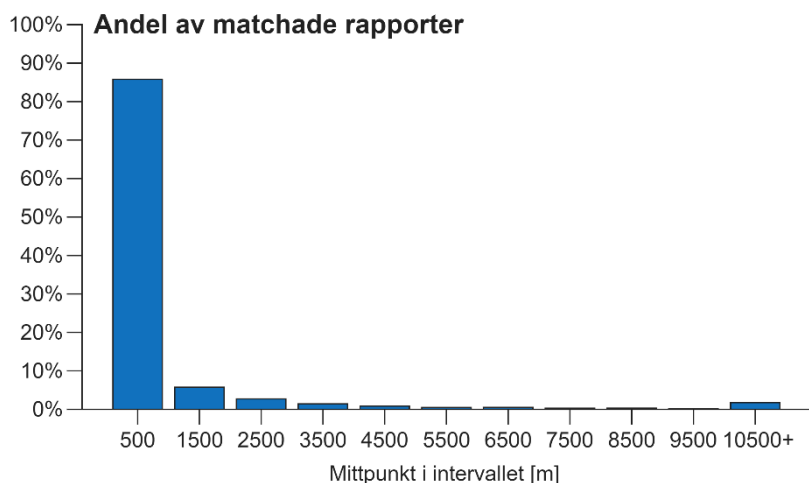
Tabell 11. Skillnad i olyckstid mellan polis- och sjukvårdsrapport i par matchade på personnivå. Baseras på 10 901 matchade par av rapporter där klockslag är angivet i båda källorna.

Timmar	Antal matchade par	Andel matchade par
0	10 328	94,7 %
1	337	3,1 %
2	50	0,5 %
3–24	186	1,7 %
Total	10 901	100,0 %



Figur 5. Fördelning över skillnad i olyckstid mellan polis- och sjukvårdsrapport i par matchade på personnivå. Baseras på 10 901 matchade par av rapporter där klockslag är angivet i båda källorna.

Uppgift om plats saknas i sjukvårdsrapporter i 1 102 av de 12 115 matchade paren. Avstånd mellan polisens och sjukvårdens platsangivelser i resterande par visas i Figur 6. I cirka 86 procent av fallen är avståndet under 1000 m och i knappt 2 procent är avståndet över en mil. Långa avstånd kan till exempel förekomma om olyckan har inträffat på en landsväg som går genom likartad omgivning och där det kan vara svårt att avgöra exakt var man befinner sig. Notera att i de fall där avståndet mellan platsangivelserna är över 1000 sker en manuell granskning av om det är samma olycka eller inte.



Figur 6. Fördelning över avstånd mellan polisrapporterad och sjukvårdsrapporterad olycksplats i rapporter matchade på personnivå. Intervallet 10500+ innehåller alla avstånd från 10 000 m och uppåt. Baseras på de 11 013 matchade par av rapporter där olycksplats är angiven från båda källorna.

I Tabell 12 visas avstånd uppdelat efter olyckstyp och vägmiljö. Generellt är det större medianavstånd utom än inom tätbebyggt område. Bäst överensstämmelse är det för olyckor mellan motorfordon och gång- eller cykeltrafikanter, där medianavståndet är 15 meter inom tätbebyggt område och 47 meter utanför tätbebyggt område.

Tabell 12. Sammanställning av avstånd mellan polisrapporterad och sjukvårdsrapporterad olycksplats i rapporter matchade på personnivå. Uppdelat efter olika typer av olyckor och vägmiljö där förkortningarna GCM (gång, cykel eller moped), GC (gång eller cykel), och MF (motorfordon) används.

Olyckstyp	Tättbebyggt område	Antal matchade par	Saknade värden	P25 (m)	Median (m)	P75 (m)	Max (m)
Alla	Alla	11 013	1 102	12	53	313	206 877
Alla	Inom	5 486	576	7	26	126	111 343
Alla	Utanför	5 224	481	24	114	730	206 877
Olyckor med bara GCM-trafikanter	Inom	785	74	11	30	95	54 134
Olyckor med bara GCM-trafikanter	Utanför	146	15	23	67	241	5 118
Olyckor mellan MF och GC	Inom	2 124	207	5	15	71	30 124
Olyckor mellan MF och GC	Utanför	207	17	9	47	229	14 633
Olyckor med bara MF	Inom	2 215	265	9	39	238	111 343
Olyckor med bara MF	Utanför	4 069	371	24	117	777	206 877

Indelning i trafikantkategorier har också jämförts mellan polis- och sjukvårdsrapporter, vilket visas i Tabell 13. Det är endast indelning i huvudkategori som jämförts, inte undergrupper. I 94,4 procent av de matchade paren har polis och sjukvård angett samma kategori. Den största avvikelserna är 394 fall där polisen angett lastbil och sjukvården angett personbil. En närmare kontroll visade att 370 av dessa var lätt lastbil.

Tabell 13 Jämförelse mellan trafikantkategori angiven i polis- och sjukvårdsrapport i par matchade på personnivå.

Trafikantkategori i sjukvårdsrapport	Trafikantkategori i polisrapport									Total
	Buss	Cykel	Fotgängare	Lastbil	Moped	Motorcykel	Personbil	Spårbundet fordon	Övrigt	
Buss	76	0	0	1	0	0	6	0	0	83
Cykel	0	2 150	46	0	10	4	3	0	10	2 223
Fotgängare	5	53	836	0	1	0	18	0	1	914
Lastbil	0	0	0	199	0	0	12	0	5	216
Moped	0	1	0	1	632	21	0	0	13	668
Motorcykel	0	0	0	0	7	762	0	0	1	770
Personbil	2	0	4	394	4	1	6 476	0	12	6 893
Spårbundet fordon	0	0	0	0	0	0	0	0	2	2
Övrigt	0	1	3	5	17	14	4	0	302	346
Total	83	2 205	889	600	671	802	6 519	0	346	12 115

Tabell 14 jämförs olyckstyp för de matchade par där både polis och sjukvård angivit personbil som trafikantkategori. Olyckstyperna överensstämmer i 75,6 procent av paren. Detta kan tyckas vara relativt lågt men vissa av olyckstyperna kan vara ganska svåra att skilja mellan, till exempel avsvängande (A) och korsande (K). När det gäller singelolyckor (S) är det bättre överensstämmelse, av

de 2 086 singelolyckor som polisen rapporterade så hade även sjukvården angett singelolycka i 93,6 procent av fallen.

Tabell 14. Jämförelse mellan olyckstyp angiven i polis- och sjukvårdsrapport i par matchade på personnivå. Baseras på de matchade par där både polis och sjukvård angivit personbil som trafikantkategori. Resterande olyckstyper inkluderar bland annat kollision mellan motorfordon och gång-cykel, vilt och varia.

Olyckstyp i polisrapport								
Olyckstyp i sjukvårdsrapport	A	K	M	O	S	U	Resterande olyckstyper	Total
A (avsvängande motorfordon)	109	160	47	19	3	74	55	467
K (korsande motorfordon)	70	503	41	5	1	35	28	683
M (möte motorfordon)	37	38	528	3	3	21	48	678
O (omkörning motorfordon)	8	4	5	32	1	29	5	84
S (singel motorfordon)	10	6	34	14	1 953	24	78	2 119
U (upphinnande motorfordon)	121	59	25	69	10	1 385	202	1 871
Resterande olyckstyper	10	16	9	6	115	32	386	574
Total	365	786	689	148	2 086	1 600	802	6 476

Eftersom det krävs fullständigt personnummer för att identifiera om en polisrapport och en sjukvårdsrapport handlar om samma person bör personer med annan nationalitet än svensk och som bara vistas tillfälligt i Sverige vara underrepresenterade i matchningen. Vi har studerat detta genom att använda variabeln *Nationalitet* som finns angivet i polisrapporter. I Tabell 15 visas tydligt att de med annan och de med okänd nationalitet har betydligt lägre andel matchade än de med svensk nationalitet.

Tabell 15. Andel matchade på personnivå av totalt antal polisrapporterade uppdelat efter nationalitet (svensk eller annan).

Nationalitet	Totalt antal polis-rapporterade personer	Antal matchade på personnivå	Andel matchade av totalt antal polisrapporterade
Svensk	38 575	11 893	30,8 %
Annan	4 658	188	4,0 %
Okänt	3 318	34	1,0 %
Total	46 551	12 115	26,0 %

För att undersöka hur viktigt personnummer är för matchningen beräknades hur många av de par som matchats på personnivå som skulle matchas till samma olycka även om personnumret inte var tillgängligt. Då baseras matchningen på tidsdifferens, avstånd mellan angivna platser och olyckstyp. Resultatet visar att totalt sett skulle 69,2 procent av de 12 115 matchade paren kopplas till samma olycka (Tabell 16). I de par där koordinater finns angivna från både polis och sjukhus skulle 76,1 procent matchas.

Tabell 16. Beskrivning av hur många av de rapporter som är matchade på personnivå som skulle matchas (då på olycksnivå) även om fullständigt personnummer inte var tillgängligt.

Koordinater	Antal matchade par på personnivå	Antal matchningar som skulle ske om personnummer ej fanns	Andel matchningar som skulle ske om personnummer ej fanns
Värden finns	11 013	8 384	76,1 %
Värden saknas	1 102	0	0,0 %
Totalt	12 115	8 384	69,2 %

4.5. Data från Nationella vägdatatabasen (NVDB)

Olyckor med koordinater angivna kopplas till information om vägnätet i NVDB. I det här avsnittet beskrivs hur den kopplade informationen är inlagd i Strada och problem som kan uppstå när den informationen ska tolkas.

I Strada prioriteras polisens angivna koordinater framför sjukvårdens för matchade olyckor. En olycka kan kopplas till flera olika väglänkar vilket innebär att väginformation i Strada kan vara tvetydig. Om flera kopplingar gjorts anges samtliga variabelvärden som en semikolonseparerad textsträng. Till exempel kan NVDB-variabeln *Vägtyp* för en specifik olycka ha värdet {Vanlig väg; ; Motorväg}. Detta betyder att olyckans koordinater kopplats till tre väglänkar där den första har värdet ”Vanlig väg”, den andra har inget värde angivet i NVDB, och den tredje väglänken har värdet ”Motorväg”. I studerade data återfinns olyckor med upp till 12 kopplade väglänkar. För varje koppling finns ett avstånd angivet i meter mellan olyckan och väglänkens kopplingspunkt. Tabell 17 anger sammanfattande statistik om antalet olyckor och kopplingsavstånd uppdelat i olika antal vägkopplingar.

Tabell 17. Sammanfattande statistik om antalet olyckor och kopplingsavstånd uppdelat i olika antal vägkopplingar för 2016–2024.

Antal vägkopplingar	Antal olyckor	Andel olyckor	Medelavstånd (m)	Maximumavstånd (m)
0	91 822	23,1 %	-	-
1	233 191	58,6 %	4,1	8 033,8
2	23 774	6,0 %	3,5	19,8
3	30 308	7,6 %	3,5	19,9
4	16 815	4,2 %	3,4	19,9
5+	1 828	0,5 %	5,1	19,8
Total	397 738	100,0 %	3,8	8 033,8

Tabell 17 visar att 23,1 procent av olyckorna har noll kopplingar och saknar därför information från NVDB. En andel av dessa beror på att vissa olyckor (10,0 procent), saknar koordinater, men vad resterande andel olyckor med saknad NVDB-information beror på kan inte utläsas ur datamaterialet. Drygt hälften av olyckorna har en enstaka vägkoppling och medelavståndet är då 4,1 meter, men maximumavståndet visar att det förekommer kopplingsavstånd upp till drygt 8 kilometer. Så långa avstånd är dock ovanliga där 99:e percentilen är 42,1 meter för dessa olyckor. För olyckor med fler än en vägkoppling visar Tabell 17 att avståndet alltid är inom 20 meter vilket tyder på att detta är en maximumgräns implementerad i Strada för olyckor med fler än en vägkoppling.

Tabell 17 visar att endast 58,6 procent av alla olyckor har en entydig vägkoppling. Notera dock att för vissa olyckor (till exempel korsningsolyckor) kan flera vägkopplingar vara relevanta, men användaren

ges ingen vägledning om detta. Ur ett användarperspektiv är det relevant att per variabel undersöka hur stor andel av alla olyckor som har fullständiga data (variabelvärde finns angivet för alla kopplingar) och samstämmiga data (alla vägkopplingar ger samma variabelvärde). Här har ett urval av variabler gjorts baserat på den information som anses mest relevant ur ett trafiksäkerhetsperspektiv. I Tabell 18 anges andelen fullständiga och samstämmiga data (FSD) per antal vägkopplingar för ett urval av relevanta variabler.

Tabell 18. Andelen fullständiga och samstämmiga data (FSD) per antal vägkopplingar för ett urval av relevanta variabler.

Antal vägkopplingar	Andel FSD Tättbebyggt område	Andel FSD Hastighetsgräns	Andel FSD Vägbredd	Andel FSD Vaghållare	Andel FSD Vägnamn	Andel FSD Vägnummer	Andel FSD Vägtyp	Andel FSD ÅDT
1	100,0 %	98,8 %	75,9 %	99,1 %	75,9 %	38,9 %	32,0 %	31,8 %
2	82,5 %	51,7 %	9,0 %	59,4 %	15,1 %	7,5 %	6,5 %	0,6 %
3	70,2 %	45,6 %	5,7 %	57,8 %	7,9 %	3,6 %	9,6 %	0,0 %
4	71,7 %	45,5 %	4,5 %	60,9 %	1,1 %	1,0 %	5,9 %	0,0 %
5+	64,6 %	34,9 %	1,6 %	55,3 %	2,2 %	1,9 %	2,3 %	0,0 %
Total	72,2 %	66,6 %	45,7 %	68,9 %	46,1 %	23,6 %	20,2 %	18,7 %

Andra raden i Tabell 18 där andelen FSD anges för olyckor med en enstaka vägkoppling ger en indikation på hur ofta information är tillgänglig i NVDB. För variablerna *Tättbebyggt område*, *Hastighetsgräns*, och *Vaghållare* finns alltid eller nästan alltid ett värde angivet. För *Vägbredd* och *Vägnamn* är andelen något lägre, 75,9 procent, och för resterande variabler *Vägnummer*, *Vägtyp*, och *ÅDT (årsdygnstrafik)* är andelen betydligt lägre, runt 32–39 procent. Notera dock att *Vägnummer* bara finns i vissa fall på kommunala och enskilda vägar vilket innebär att en låg andel FSD för denna variabel inte nödvändigtvis innebär en brist i NVDB. Även variabeln *ÅDT* finns i huvudsak bara för statliga vägar vilket förklarar den låga andelen FSD. Tabell 18 visar också att andelen FSD minskar kraftigt desto fler vägkopplingar som olyckan har, vilket visar att data ofta inte är samstämmiga för en specifik olycka. På totalen har *Tättbebyggt område* högst andel med 72,2 procent medan *ÅDT* har lägst med 18,7 procent.

Olyckans bebyggelseyp finns angiven som separat polis- och sjukvårdsvariabel och syftar till att ge samma information som NVDB-variabeln *Tättbebyggt område*; om olycksplatsen är inom tättbebyggt område eller ej. Detta möjliggör en jämförelse mellan polisens och/eller sjukvårdens angivna värde och värdet från NVDB. I Strada prioriteras polisens angivna värde för denna variabel framför sjukvårdens om olyckan har matchats. Av de totalt 397 738 olyckorna som studeras är olyckans bebyggelseyp angiven till "Tättbebyggt område" eller "Ej tättbebyggt område" för 361 739 av olyckorna (okänt värde för övriga). Av dessa har 275 951 olyckor FSD från NVDB. Resultaten visar att av dessa 275 951 olyckor stämmer polisens och/eller sjukvårdens rapportering överens med NVDB-kopplingen i 244 101 (88,5 procent) av fallen. Ett alternativ vid jämförelse mellan polisens rapportering av tättbebyggt område och NVDB är att använda den geografiskt närmsta vägkopplingen i NVDB enligt angivet kopplingsavstånd. Av de 293 935 olyckor där data finns från båda källorna överensstämde angivelsen "tättbebyggt område" eller "Ej tättbebyggt område" i 260 269 (88,6 procent) av fallen, alltså en likvärd andel som första alternativet. Sammantaget innebär det att överensstämmelsen mellan polisens och/eller sjukvårdens rapportering av bebyggelseyp och NVDB-data är relativt god i de fall där data är angivna från båda källorna.

Polisrapporterad hastighetsbegränsning kan också jämföras med angiven hastighetsbegränsning i NVDB. För att möjliggöra jämförelse begränsas här analysen till olyckor där endast väg A har en angiven hastighetsbegränsning (olyckor med angiven hastighetsbegränsning på väg B som används vid korsningsolyckor sällas alltså bort). Detta ger totalt 73 292 olyckor, varav 63 740 med FSD från NVDB. Av dessa 63 740 olyckor överensstämde angiven hastighetsbegränsning i 50 969 (80,0 procent) av fallen. Om närmsta vägkoppling i NVDB används istället så var överensstämmelsen lägre, 54 776 av 90 342 (60,6 procent) av fallen hade då samma värde.

Som redovisat i avsnitt 4.2 slutade polisen ange vägtyp för olyckan runt 2020. Detta innebär att NVDB-data nu är enda källa i Strada för denna variabel. Även data innan 2020 där både polis- och NVDB-rapporterade data finns är svåra att jämföra eftersom variabelernas värdemängder skiljer sig åt. Polisen angav vägtyp enligt:

- ”Annan allmän väg”
- ”Tom”
- ”Enskild väg”
- ”Gata”
- ”Motortrafikled”
- ”Motorväg”
- ”Okänt”
- ”Uppgift saknas”
- ”Övrig väg”

I NVDB anges däremot följande värden:

- ”4-fältsväg”
- ”Tom”
- ”Motortrafikled”
- ”Motortrafikled mötesfri”
- ”Motorväg”
- ”Vanlig väg”
- ”Vanlig väg mötesfri”

Motorväg och motortrafikled anges dock från båda källorna. Med samma metodik som ovan visar resultaten att av de 5 493 olyckor där polisen angivit vägtyp ”Motorväg” och det fanns FSD NVDB-data hade 4 718 (85,9 procent) samma värde i NVDB. Motsvarande antal när närmsta NVDB-koppling användes var 5 596 respektive 4 756, vilket innebär en överensstämmelse på 85,0 procent. För motortrafikled var antalet olyckor ungefär hälften och andelarna betydligt lägre, 11,5 procent både för FSD och närmaste kopplingen.

4.6. Dokumentation och metadata

Data som ligger till grund för analyserna i avsnitt 4.1–4.5 har hämtats från Strada uttagswebb. I detta avsnitt presenteras resultat och insikter gällande datauttagens struktur och behovet av dokumentation ur ett användarperspektiv.

4.6.1. Övergripande dokumentation

På Transportstyrelsens hemsida för Strada⁶ finns en användarhandledning för Strada uttagswebb (Transportstyrelsen, 2026). Det framgår inte från handledningen vilket datum den publicerats. Denna handledning innehåller en kort inledning om syftet med Strada, hur användaren gör sitt dataurval, uttagsstruktur, listor med variabelnamn, samt vissa begreppsförklaringar. En lista med sjukhus med år och månad när de anslöts till Strada ingår också som bilaga. Ur ett användarperspektiv är ett sammanhållet dokument av denna typ värdefullt (även för de som inte får data via uttagswebben), men det skulle kunna kompletteras med mer information. Ett förslag är att presentera en tidslinje med större förändringar som kan påverka datakvalitet och bortfall. Detta skulle till exempel kunna inkludera lagförändringar, rutinförändringar, samt uppdateringar, tillkomst och borttagning av variabler. I vissa fall finns rapporter som utvärderar förändringarna och då skulle länkar till dessa kunna finnas direkt i tidslinjen.

Dödsfall och sjukvårdens skadegrader är beskrivna, men kan kompletteras med polisens definitioner av lindrig och svår skada. Kopplingen till NVDB beskrivs kortfattat, men logiken kring hur kopplingen utförs och dessa data presenteras är inte beskriven. Vidare är Stradas koppling till Vägtrafikregistret för körkortsinformation i nuläget inte beskriven i dokumentet.

4.6.2. Tabellstruktur för datauttag

Datauttag laddas ner från Strada uttagswebb i MS Excel-format (.xlsx). Om användaren valt att inkludera alla tillgängliga data innehåller Excel-filen sex flikar: en flik med sammanfattande information och fem flikar med varsin datatabell med namn *Olyckor*, *Personer*, *Personer polis*, *Personer sjukvård*, och *Skador*. Samtliga fem tabeller innehåller olycksnummer och de fyra sistnämnda tabellerna innehåller också ett personspecifikt referensnummer. Dessa möjliggör länkning av dataposter mellan tabeller. Excelformatet innebär att användaren direkt kan se data och göra enklare analyser, men det kan också innebära vissa hinder för användaren. Excel har stor spridning och användning, men har ett begränsat antal rader per flik och är inte optimerat för stora datamängder. Detta innebär att användaren kan behöva dela upp ett större datauttag i flera mindre uttag för att få ut filer som går att arbeta med. En lösning på detta problem är att erbjuda nedladdning av datatabeller i ett enkelt textformat (till exempel CSV). Detta skulle förenkla inläsning i statistikmjukvaror såsom SAS, SPSS, Matlab, R och Pythonbaserad mjukvara.

Information på personnivå finns i tre tabeller med data rapporterade från polisen (*Personer polis*), från sjukvården (*Personer sjukvård*), och med alla personer från en eller båda källorna (*Personer*). *Personer* innehåller både källspecifika variabler och variabler som rapporteras av båda källorna. De sistnämnda prioriteras eller sammanvägs till ett värde om båda källor finns för olyckan. Enligt Transportstyrelsen prioriteras polisens uppgifter gällande olyckan och sjukvårdens uppgifter gällande skador.

Information på olycksnivå finns enbart i en tabell, *Olyckor*, med prioriterad och sammanvägd information på liknande sätt som för tabellen *Personer*. Det finns inte separata tabeller för olyckor polis och olyckor sjukvård som det finns för personer. Troligtvis är detta på grund av att sjukvården inte rapporterar på olycksnivå och hur matchningsalgoritmen skapar olyckor.

Tabellen *Olyckor* innehåller också variabler med de trafikelement som var inblandade i olyckan. Tyvärr saknar dessa koppling till de personer som var inblandade i olyckan. Detta innebär en

⁶ <https://www.transportstyrelsen.se/STRADA>

begränsning, till exempel när flera fordon av samma typ är involverade i olyckan, då data inte visar vem som satt i vilket fordon⁷.

Sammantaget är tabellstrukturen med uppdelning i olyckor, personer, och skador logisk, men tydligare metadata och variabelbeskrivningar behövs för att förstå vad varje variabel innebär. Detta beskrivs i nästkommande stycke.

4.6.3. Variabelspecifik metadata

På Transportstyrelsens hemsida finns en variabellista i Excelformat som innehåller namn, datatyper, beskrivningar, värdemängder och rapporteringskällor. I filen finns det en flik för olycksnivå och en för personnivå. Det finns också en tredje flik som innehåller värdemängder för vissa specifika variabler. Figur 7 och Figur 8 visar exempelbilder på hur strukturen i tabellerna ser ut i filen.

	A	B	C	D	E	F	G	H	I	J
1	Uppdaterad: 2023-01-23		Gemensam data, polis & sjukvård	Polisdata	Sjukvårdsdata	NVDB	VTR (för deras värdemängder se Transportstyrelsens hemsida)			
2										
3	Personnivå									
4	Variabel i uttagswebben	Antal variabel (Datatyp)	Beskrivning	Defaultvärde	Värdeområde (se filen Väg- värdeområden)	Polis och/eller Sjukvård	Källa	fr o m - to m (se Report.DateOfRegist ration om inget annat anges)	kommentar	Klicka på rutaorna för att markera variabler
5	Olycksnummer	Int	Olycksnummer (samma som Olycksid i gamla Strada)			P + S	Programgenererad			<input type="checkbox"/>
6	Referens	Int	Personreference ID		References : strada.PersonReference	P + S	Programgenererad			<input type="checkbox"/>
7	Polisrapport	Int	Antal polisrapporter		Ja; Nej	P	Programgenererad			<input type="checkbox"/>
8	Sjukvårdsrapport	Int	Antal sjukhusrapporter		Ja; Nej	S	Programgenererad			<input type="checkbox"/>
9	År	Datum	År från olycksdatum		2000-	P + S	Programgenererad utifrån olycksdatum			<input type="checkbox"/>
10	Månad	Datum	Månad från olycksdatum		1-12	P + S	Programgenererad utifrån olycksdatum			<input type="checkbox"/>
11	Dag	Datum	Dag från olycksdatum		1-31	P + S	Programgenererad utifrån olycksdatum			<input type="checkbox"/>
12	Veckodag	Int	Veckodag		Måndag-Söndag	P + S	Programgenererad			<input type="checkbox"/>
13	Osäkert datum	Bit	Osäkert datum total			P + S				<input type="checkbox"/>
14	Okänt datum	Bit	Datum okänt, enbart fr sjukvården.		0 - 1	S				<input type="checkbox"/>
15	Klockslag grupp (timme)	Int	Klockslag (enbart timme)		00-24	P + S				<input type="checkbox"/>
16	Klockslag	Tid	Klockslag då olyckan inträffade		00:00-23:59	P + S				<input type="checkbox"/>
17	Osäkert klockslag	Bit	Indikator om olycksfallstidpunkten är osäker		Ja eller tom	P + S				<input type="checkbox"/>
18	Okänt klockslag	Bit	Indikator om olycksfallstidpunkten är okänd		Ja eller tom	P + S				<input type="checkbox"/>
19	Län	Int	Län		ValueCounty	P + S				<input type="checkbox"/>
20	Kommun	Int	Ort/Kommun		ValueMunicipality	P + S				<input type="checkbox"/>
21	Bebyggelseyp	Int	Bebyggelseyp		Tättbebyggt område; Ej tättbebyggt område;	P + S				<input type="checkbox"/>
22	Olycksväg/-gata	sträng	Namn eller nummer på den väg som är en del av länksekvensen			P + S	Programgenererad fr NVDB			<input type="checkbox"/>
23	Olycksplats	sträng	Beskrivning av olycksplatsen			P + S				<input type="checkbox"/>
24	Position i kartan	Bit	Indikator om position är osäker eller okänt		Obestämbar position; Osäker position; tom	P + S				<input type="checkbox"/>
25	Platstyp	Int	Olycksplatstyp		ValuePlaceType	P + S				<input type="checkbox"/>
26	Attribut till platstyp	Int	Attribut till olycksplats		ValuePlaceAttribute	P + S	Programgenererad			<input type="checkbox"/>
27	Olyckstyp	Int	Olycksfallstyp		ValueAccidentTypeGroup	P + S				<input type="checkbox"/>
28	Händelseförlopp (P)	sträng	Beskrivning av olyckshändelsen enl polisen			P				<input type="checkbox"/>
29	Händelseförlopp (S)	sträng	Beskrivning av olyckshändelsen enl sjukvården (patienten)			S				<input type="checkbox"/>

Figur 7. Bild från fliken Personnivå i Stradas variabelista.

⁷ Enligt Transportstyrelsen har denna information lagts till så att det nu går att koppla trafikelement till person i datauttaget. Detta har dock inte varit fallet i tidigare datauttag eller datauttaget som användes i aktuell studie. Inom denna studie har vi inte kunnat verifiera att det nu finns en koppling för samtliga trafikelement och år.

ValueAccidentOccurredWhen										ValueAccidentSeverity												
ID	CreatedAt	Updated	Deleted	SortOrder	ValidFrom	ValidTo	ValueAccidentOccurredWh	Hidden	Journal	ID	CreatedAt	Updated	Deleted	SortOrder	ValidFrom	ValidTo	ValueAccidentSeverity	Priority	PersonValue	PoliceAc	PolicePersonValue	
1	2020-06-02		FALSE	9	1997-11-01		Okänt		TRUE	1	2020-06-02		FALSE	1	1753-01-01		Dödsolycka	1	Död			
2	2020-06-02		FALSE	1	1997-11-01		I arbetet		FALSE	2	2020-06-02		FALSE	2	1753-01-01		Måttlig olycka (ISS 4-8)	13	Måttligt skadad (ISS	Svår	Svårt skadad	
3	2020-06-02		FALSE	2	1997-11-01		På väg till/från arbetet		FALSE	3	2020-06-02		FALSE	3	1753-01-01		Lindrig olycka (ISS 1-3)	14	Lindrigt skadad (ISS 1-	Lindrig	Lindrigt skadad	
4	2020-06-02		FALSE	4	1997-11-01		I skolan		FALSE	4	2020-06-02		FALSE	4	1753-01-01		Ej personskadeolycka	17	Oskadad			
5	2020-06-02		FALSE	5	1998-04-01		På väg till/från skolan		FALSE	5	2020-06-02		FALSE	5	1753-01-01		Allvarlig olycka (ISS 9-)	12	Allvarligt skadad (ISS			
6	2020-06-02		FALSE	7	1997-11-01		På fritiden		FALSE	6	2020-06-02		FALSE	6	1753-01-01		Olycka med osäkra skador	16	Osäker skada			
7	2020-06-02		FALSE	8	1999-01-01		Annat		TRUE	7	2020-06-02		FALSE	7	1753-01-01		Olycka med osäkra skador	15	Osäker skada	Okänd	Okänd skadegrad	
8	2020-06-02		FALSE	8	2020-06-02				FALSE	8	2020-06-02		FALSE	8	1753-01-01		Dödsolycka, ej officiell	2	Död, ej officiell			
9	2020-06-02		FALSE	9	2020-06-02				FALSE	9	2020-06-02		FALSE	9	1753-01-01		Dödsolycka, ej officiell	3	Död, ej officiell			
10	2020-06-02		FALSE	10	2020-06-02				FALSE	10	2020-06-02		FALSE	10	1753-01-01		Dödsolycka, ej officiell	4	Död, ej officiell			
11	2020-06-02		FALSE	11	2020-06-02				FALSE	11	2020-06-02		FALSE	11	1753-01-01		Dödsolycka, ej officiell	5	Död, ej officiell			
12	2020-06-02		FALSE	12	2020-06-02				FALSE	12	2020-06-02		FALSE	12	1753-01-01		Dödsolycka, ej officiell	6	Död, ej officiell			
13	2020-06-02		FALSE	13	2020-06-02				FALSE	13	2020-06-02		FALSE	13	1753-01-01		Dödsolycka, ej officiell	7	Död, ej officiell			
14	2020-06-02		FALSE	14	2020-06-02				FALSE	14	2020-06-02		FALSE	14	1753-01-01		Dödsolycka, ej officiell	8	Död, ej officiell			
15	2020-06-02		FALSE	15	2020-06-02				FALSE	15	2020-06-02		FALSE	15	1753-01-01		Dödsolycka, ej officiell	9	Död, ej officiell			
16	2020-06-02		FALSE	16	2020-06-02				FALSE	16	2020-06-02		FALSE	16	1753-01-01		Dödsolycka, ej officiell	10	Död, ej officiell			
17	2020-06-02		FALSE	17	2020-06-02				FALSE	17	2020-06-02		FALSE	17	1753-01-01		Dödsolycka, ej officiell	11	Död, ej officiell			

Figur 8. Bild från filen Väg-värde mängder i Stradas variabellista.

Med utgångspunkt från att dokumentation ska vara korrekt och lättförståelig för användaren, samt att metadata ska vara maskinläsbara (alltså följa ett format som kan läsas in av olika mjukvaror på ett standardiserat sätt), visar genomgång av filen på flera brister. Många variabler i Strada är kategoriska textsträngar i datauttaget som laddas ner från Stradas uttagswebb, men flertalet av dessa har angiven datatyp "Int" (heltal). Vidare är värde mängder inte beskrivna i ett enhetligt format och vissa är inte fullständiga. Till exempel är värden ibland semi-kolonseparerade, ibland separerade med ordet "eller", samt ibland separat redovisade i en tabell i filen Väg-värde mängder. Vidare anges inte intervall för numeriska värden enhetligt. Väg-värde mängder filen är heller inte anpassad för snabb maskinell inläsning eftersom den består av flera tabeller uppräddade efter varandra istället för att vara organiserade i en tabell. Det vore värdefullt med information om när variabler infördes och när förändringar gjordes. För vissa variabler går detta att se indirekt i filen Väg-värde mängder, men vissa värden är uppenbart inkorrekta. Till exempel har variablerna Alkoholpåverkan och Drogpåverkan samma värde mängd "ValueIntoxication" som är angiven att gälla från 1753. Som visades i Tabell 6 och Tabell 7 verkar dock variabeln Drogpåverkan ha införts år 2020 liksom svarsalternativet "Misstanke". Exceldokumentet saknar också en sammanfattande flik som beskriver vad kolumnerna i de olika tabellerna står för.

Som redovisat i 4.2 är det oklart vad "Tom" (tomma celler i Excel) betyder och här förslås en utveckling av värde mängderna där "Tom" till exempel ersätts med "Nej", "Uppgift saknas", "Okänt", "Ej i bruk" och "Ej relevant" beroende på betydelsen. När data saknas för att det inte finns en rapport från den källan skulle detta till exempel kunna anges som "Ej polis" eller "Ej sjukvård". Samma uppdelning för numeriska variabler är svårare att hantera i Excel eftersom "Tom" är det enda tillgängliga flaggan för saknade värden. Om man däremot erbjöd datauttag i textformat skulle man kunna använda olika benämningar på saknade numeriska data såsom "NAa" och "NAb".

5. Diskussion

Problemet med minskad rapportering från polisen har tidigare belysts av Transportstyrelsen (2024). Våra resultat visar att den minskade rapporteringen också gäller dödsfall. För 2016–2024 ses en negativ trend där flest dödsfall (34 personer) saknades 2024. De saknade dödsfallen beror på att polisen inte längre registrerar alla dödsfall som kommer till deras kännedom i efterhand, att polisen inte heller registrerar alla som dör på sjukhus till följd av olyckan, samt att personer med skyddade personuppgifter inte registreras i Strada. Att Strada inte är heltäckande gällande dödsfall får negativa konsekvenser för dess användare. På kommunal nivå blir det svårt att följa utvecklingen av antalet dödade om vissa fall saknas då antalet dödade redan kan vara få för specifika kommuner. På statlig nivå innebär det svårigheter att genomföra heltäckande analyser och statistiksammanställningar för myndigheter såsom Trafikverket, Trafikanalys, och VTI. Till exempel är VTI som myndighet och forskningsinstitut beroende av Strada i sin forskning och för att leverera statistik till internationella organisationer såsom European Transport Safety Council (ETSC) och Nordiskt vägforum (NVF). Detta kräver oftast andra indelningar gällande till exempel ålder, olyckstyp och vägtyp, än vad som är tillgängliga i den officiella statistiken från Trafikanalys. Det leder i sin tur till att inrapporteringen blir ofullständig eller att uppskattningar används vilket inte borde behövas vad gäller dödsfall. Ett alternativ är att Transportstyrelsen tar fram de önskade indelningarna vilket leder till mer jobb för dem. Transportstyrelsen (2025) anger i sin färdplan för Strada att polisens minskade rapportering och bortsällningen av personer med skyddade personuppgifter är prioriterade problem att åtgärda. Förhoppningsvis kan dessa problem lösas genom dialog mellan berörda myndigheter och departement, eller en uppgraderad säkerhetsklassning av Strada.

Vår genomgång av partiellt bortfall inom polis- och sjukvårdsrapporter visar på svårigheter att förstå och tolka vissa variabelvärden. Detta beror främst på avsaknad av dokumentation och metadata där bland annat värdemängderna inte i tillräckligt hög utsträckning skiljer på olika typer av saknade data. Detta försvårar också systematisk uppföljning av datakvaliteten. Partiellt bortfall skulle kunna följas över tid som ett kvalitetsmått om olika värden angavs för fallen: (1) källan har beaktat variabeln i sin rapportering och svaret är nej, (2) källan har beaktat variabeln och svaret är okänt, (3) källan har missat att rapportera värdet för denna variabel, (4) variabeln är inte relevant för aktuell olycka eller (5) variabeln var ej i bruk vid den aktuella tidpunkten. Det går ändå att utläsa från resultaten att polisrapporter i princip alltid innehåller datum och tid, position, olyckstyp, olycksplats, och en beskrivning av händelseförloppet. Data visar också att det partiella bortfallet gällande trafikregler och olycksomständigheter minskar årligen. Detta skulle kunna påverka forskningen om effektsamband mellan olyckor och dessa faktorer positivt, även om förändringar över tid måste hanteras metodmässigt. För sjukvårdsrapporter finns i princip alltid datum, olyckstyp, trafikantkategori, kön, och ålder angivet, men saknade data för position och klockslag har ökat de senaste åren och ligger idag på nivåer kring 20, respektive 15 procent. Detta pekas också ut av Transportstyrelsen (2025) som problematiskt. Ökningen startade i samband med lagförändringen men detta behöver studeras närmare för att förstå vad den beror på. I motsats till polisen visar sjukvårdsdata också en ökning av det partiella bortfallet gällande olycksomständigheter, vilket skulle kunna få negativa konsekvenser för forskningen och utvärdering av trafiksäkerhetsåtgärder. Kvaliteten av variablerna för alkohol- och drogpåverkan kan också ifrågasättas då andelen saknade värden skiftar stort mellan år.

Resultaten visar att andelen polisrapporter som saknar diarienummer ökat markant från 2020, där 56 procent saknades 2024. Saknade diarienummer från polisen gör det svårare att koppla ihop Strada med polisutredningar för fördjupade analyser. Enligt Transportstyrelsen har polisen beslutat att diarienummer inte ska ingå när rapporter skickas in via TRAFÖ. Dessa rapporter får i stället ett TRAFÖ-nummer. Detta nummer finns dock inte med i uttag från uttagswebben och det är oklart hur kopplingen kan göras mellan polisens rapporter i Strada och polisens egna utredningar.

Strada är unik genom användandet av två källor, där liknande databaser i omvärlden endast är baserade på polisrapporterade data. Att matchningen fungerar bra är viktigt eftersom missade

matchningar leder till dubbelräkning av personer och felaktiga matchningar leder till sammanräkningsfel och att fel information kopplas till olyckan. Ur den aspekten är det anmärkningsvärt att det saknas fullständig dokumentation tillgänglig om hur matchningen av polis- och sjukvårdsrapporter går till, detta gäller speciellt i de fall där fullständiga personnummer saknas. Dokumentationen som erhöles från Transportstyrelsen angående matchningen var i vissa delar förlegad och ofullständig och behövde uppdateras med hjälp av Transportstyrelsens IT-förvaltning. Beskrivningen som presenteras i denna rapport får ses som ett första steg att dokumentera och sprida information om hur matchningen går till. Heltäckande utvärdering av matchningens funktion kräver både anseilig tid och tillgång till interna data och är därför utanför ramen för denna studie. Personnummer är centrala för matchningen med ytterligare kontroller gällande tid och avstånd. Detta får anses vara ett rimligt tillvägagångssätt även om det är svårt att utläsa motivet till de relativt komplexa beräkningarna av Q-värden. Det är också oklart hur man kommit fram till ekvationernas utseende och konstanternas värden.

Resultaten baserade på data från 2022–2023 (exklusive fallolyckor) visar att 12 115 av 93 427 personer matchats på personnivå, vilket motsvarar 13,0 procent. Denna förhållandevist låga andel behöver inte vara en indikation på att matchningen fungerar dåligt, utan beror troligtvis på att täckningen delvis skiljer sig åt mellan källorna. Det är till exempel känt sedan tidigare att sjukvården fångar betydligt fler olyckor med cyklister än vad polisen gör (Amin et al. 2022). För specifika matchningsgrader och djupare analys av detta hänvisas till Forsman och Howard (2026). Sett till matchade personpar skiljer sig rapporterad tid och position ganska lite mellan polisen och sjukvården. Resultaten visar att 94,7 procent har en tidsskillnad på mindre än en timme och att 85,8 procent har en avståndsskillnad under 500 meter. Detta indikerar att gränserna på 24 timmar och 1 000 meter (där manuell kontroll sker vid överskridande) inte är för restriktiva. Gällande avståndsskillnaden visar resultaten också att den generellt är mindre inom tätbebyggt område, och för olyckor mellan motorfordon och gång- eller cykeltrafikanter. En förklaring till detta kan vara att det är enklare att bestämma exakt olycksposition i trafikmiljöer med korsande vägar och landmärken att förhålla sig till, jämfört med landsvägar där trafikmiljön kan vara mer homogen. Parkeringsmöjligheter kopplat till området skulle också kunna påverka detta beroende på om polisen anger koordinater vid sin bil eller exakt vid olycksplatsen.

Genomgången av vägkopplingar och variabler från NVDB visar att data i sin nuvarande form har låg användbarhet. Denna slutsats beror på att: (1) flera väglänkar ofta kopplas till olyckan och det är svårt för användaren att avgöra vilka som är relevanta, (2) kopplingarna är strukturerade som en textsträng i stället för som separata variabler, och (3) vissa data saknas i NVDB eller finns bara för statligt vägnät. En vägledning till Stradanvändaren som beskriver hur olyckor har kopplats och hur kopplingen kan användas vore därför värdefullt. Att polisen har slutat att rapportera vägtyp är olyckligt då våra resultat visar att endast 20,2 procent av olyckorna har fullständiga (variabelvärde finns angivet för alla kopplingar) och samstämmiga (alla vägkopplingar ger samma variabelvärde) data gällande vägtyp från NVDB. Vägtyp är en av de variabler som frekvent efterfrågas vid statistiska sammanställningar både nationellt och internationellt. I den officiella statistiken som publiceras av Trafikanalys⁸ sker idag dubbelräkning där Trafikanalys anmärker att ”När det gäller vägtyp, det förekommer olyckor där flera väglänkar är kopplade till olyckspositionen och i dessa fall räknas unika vägtyper. Detta gör att tabellen summerar till fler olyckor och personer än totalen.” Här skulle man kunna undersöka möjligheten att återinföra vägtyp som polisvariabel, förslagsvis med samma värdemängd som i NVDB. Möjligheter för rapportören att välja exakt väg i sitt inrapporteringsverktyg bör också undersökas. Det ska dock tilläggas att aktuell analys inte studerat kopplingen i detalj gällande uppdelning i till exempel polis- och sjukvårdsrapporterade data eller uppdelat för olika trafiktyper. Vissa olyckor ska ju också ha mer än en vägkoppling om de sker i en korsning.

⁸ <https://www.trafa.se/vagtrafik/vagtrafikskador/>

Analyserna och resultaten visar på ett behov att utöka och förbättra Stradas dokumentation och metadata, där saknade variabelvärden och matchningen redan nämnts ovan. Genomgång av nuvarande dokumentation visade på ett behov att tydliggöra förändringar som påverkar Stradadata, både på övergripande nivå gällande lagstiftning och rutiner, och på variabelnivå gällande när nya variabler införs eller befintliga variabler ändras. På övergripande nivå skulle en tidslinje som kartlägger större förändringar förbättra användarens förståelse för vilka tidsperioder som är relevanta vid jämförande analyser. På variabelnivå visar analyserna till exempel att förändringar av hur personer som är oskadade, har okänd skada, eller har osäker skada AIS-kodas i systemet lämnats till användaren att upptäcka själv, och det finns många andra exempel på saknad information. Här behövs både ett dokument som ger den övergripande beskrivningen av Strada och tillhörda variabelspecifika metadata.

Nuvarande användarhandledning är riktad till användare av Strada uttagswebb men innehåller också information som är relevant för de som beställer data via blankett. På samma sätt verkar variabellistan i Excelformat användas som både databeställningslista och metadatafil. En tydligare uppdelning skulle kunna förenkla för användaren och underlätta att hålla dokumentation uppdaterad. Metadata bör vara maskinläsbara (alltså följa ett format som kan läsas in av olika mjukvaror på ett standardiserat sätt) och det vore värdefullt att de följde med automatiskt till varje databeställning/uttag för att säkerställa att användaren har den senaste versionen. Här önskas också att datauttag görs tillgängliga i textformat som alternativ till nuvarande Excelformat. Gällande struktur på metadata så leder Trafikverket ett arbete med att utforma en grunddatadomän för transportsystemet där Transportstyrelsen deltar (Calles et al., 2022). Att metadata följer de strukturer som tas fram inom det projektet bör underlätta implementeringen och skapa positiva synergier med det pågående arbetet.

6. Slutsatser och rekommendationer

Syftet med denna studie har varit att beskriva Stradas datakvalitet ur ett användarperspektiv och att ta fram rekommendationer för att förbättra Stradas datakvalitet och användarbarhet. Resultaten visar att täckningen gällande dödade inte är fullgod, användarbarheten av kopplade NVDB-data är låg, och att det finns ett stort behov av bättre dokumentation. Sett till partiellt bortfall indikerar resultaten att kvaliteten ökat för polisrapporter och minskat för sjukvårdsrapporter över tid. Vidare har en beskrivning av matchningsalgoritmen tagits fram, men fullständig korrekthet kan inte garanteras givet bristande dokumentation.

Utifrån beskrivningen och resultaten i denna rapport rekommenderas Transportstyrelsen att:

1. Föra dialog med departement och de berörda myndigheterna Polisen, Trafikverket, Trafikanalys och VTI om konsekvenserna av polisens nedprioritering av rapportering till Strada.
2. Fortsätta att prioritera inkludering av information om personer med skyddade personuppgifter i Strada utan att begränsa tillgången för användarna.
3. Utöka värdemängderna för Stradas variabler för att möjliggöra tydligare indelning av saknade värden.
4. Säkerställa att Stradadata fortfarande kan kopplas till polisens interna utredningsmaterial trots minskningen av angivna diarienummer i Strada.
5. Kartlägga och dokumentera matchningsalgoritmen i detalj för att möjliggöra utvärdering och identifiering av förbättringar.
6. Utveckla en vägledning till användaren som beskriver hur olyckor har kopplats till NVDB och hur kopplingen kan användas.
7. Återinföra vägtyp som polisrapporterad variabel med samma värdemängd som i NVDB.
8. Undersöka möjligheten att utveckla polisens och sjukvårdens rapporteringsverktyg så att endast väglänkar relevanta för olyckan kopplas från NVDB, samt att strukturera dessa data i separata variabler.
9. Dela upp nuvarande användarhandledning för Strada uttagswebb i två dokument där det första innehåller övergripande information, registrets syfte, tidslinje och definitioner, och det andra innehåller information om själva uttagswebben.
10. Skapa en maskinläsbar variabelmetadatafil innehållande namn, typer, värdemängder, startdatum, slutdatum, och förklaringar. Denna bör bifogas till varje datauttag.
11. Erbjud data i textformat som alternativ till Excelformat.

Avslutningsvis ska denna rapport inte ses som en fullständig genomlysning av Stradas datakvalitet. Studien inkluderar inte jämförelse med andra register, studerar inte innehållet i varje variabel i detalj, samt adresserar inte alla datamängder (till exempel data kopplade från Vägtrafikregistret). Datakvalitet är ett vitt begrepp och denna studie har fokuserat på vissa specifika aspekter med målet att öka användarbarheten av data från Strada.

Referenser

- Amin, K., Skyving, M., Bonander, C., Krafft, M., & Nilson, F. (2022). Fall- and collision-related injuries among pedestrians in road traffic environment – A Swedish national register-based study. *Journal of Safety Research*, 81, 153–65. <https://doi.org/10.1016/j.jsr.2022.02.007>.
- Bengtsson, K. (2017). En jämförelse mellan Strada och PAR 2012: Vilken bild av antalet skadade i vägtrafiken ger de båda registren? Transportstyrelsen. <https://www.transportstyrelsen.se/globalassets/global/publikationer-och-rapporter/vag/strada/en-jamforelse-mellan-strada-och-par---tsv-2017-3763.pdf>.
- Björketun, U. (2006). STRADA, koppling mellan polis- och sjukvårdsrapporterade data. VTI notat 45-2005. VTI. <https://urn.kb.se/resolve?urn=urn:nbn:se:vti:diva-1579>.
- Calles, B.-M., Petersson, J., Lundberg, J., Slavotic, M., Andersson, U., Petterson, H., & Jansson, J. (2022). Grunddatadomän transportsystem : Samverkansuppdrag mellan transportmyndigheterna för att bedriva en utforskande utveckling att utreda en grunddatadomän för information kopplat till transportsystemet (Nr TRV 2021/56379). <https://urn.kb.se/resolve?urn=urn:nbn:se:trafikverket:diva-5316>.
- Forsman, Å. & Eriksson, O. (2023) Bortfallsuppräknings Strada: underlag till metod som baseras på både polis- och sjukvårdsrapporter. VTI PM. VTI. <https://urn.kb.se/resolve?urn=urn:nbn:se:vti:diva-19559>.
- Forsman, Å. & Howard, C. (2026). What factors affect the match rates of police and emergency department data in Sweden’s road traffic crash database Strada? Manus inskickat för publicering.
- Fredlund, T. (2016). Minskad sjukvårdsregistrering i Strada år 2015: Möjliga orsaker, konsekvenser och åtgärder. Transportstyrelsen rapport Dnr TSV 2016-4905. Transportstyrelsen.
- Fredlund, T. & Frank, M. (2025). Strada efter lag (2021:319): förändringar i rapporteringsmönster och datakvalitet. Transportstyrelsen. <https://www.transportstyrelsen.se/globalassets/global/publikationer-och-rapporter/rapporter/vagtrafik/strada-efter-lag-2021-319.pdf>.
- Hurtig, P. (2024). Analys av trafiksäkerhetsutvecklingen 2023 : Målstyrning av trafiksäkerhetsarbetet mot etappmålen 2030. Trafikverket. Hämtad från Trafikverkets hemsida: <https://urn.kb.se/resolve?urn=urn:nbn:se:trafikverket:diva-16002>.
- Transportstyrelsen (2024). Polisens rapportering till Strada – En genomgång. Opublicerat PM. Transportstyrelsen.
- Transportstyrelsen (2025). Färdplan för Strada – 2025–2031. Opublicerat dokument. Transportstyrelsen.
- Yamazaki, R. (2018). Strada bortfallshandbok 2018: Information om täckning och bortfall i rapportering till Transportstyrelsens vägolycksdatabas. Transportstyrelsen.

Bilaga 1: Andel saknade värden polis

Tabell B1.1. Andel ”Okänt”, ”Okänd”, ”Uppgift saknas”, och ”Tom” per olycksvariabel och år.

Variabel	2016	2017	2018	2019	2020	2021	2022	2023	2024
Attribut till platstyp	89 %	87 %	87 %	86 %	84 %	85 %	83 %	83 %	82 %
Bebyggelsestyp	8 %	9 %	8 %	8 %	7 %	6 %	4 %	4 %	4 %
Dag	0 %	0 %	0 %	0 %	0 %	0 %	0 %	0 %	0 %
Diariernr	0 %	0 %	0 %	0 %	2 %	19 %	45 %	54 %	56 %
E-sweref	0 %	0 %	0 %	0 %	0 %	0 %	0 %	0 %	0 %
Gatubelysning väg A	74 %	73 %	74 %	73 %	72 %	72 %	71 %	70 %	69 %
Gatubelysning väg B	95 %	95 %	95 %	96 %	95 %	96 %	96 %	95 %	95 %
Hastighetsbegränsning väg A	28 %	28 %	27 %	26 %	25 %	20 %	16 %	14 %	13 %
Hastighetsbegränsning väg B	87 %	88 %	87 %	87 %	87 %	87 %	84 %	84 %	83 %
Händelseförlopp	0 %	0 %	0 %	0 %	0 %	0 %	0 %	0 %	0 %
Klockslag	1 %	1 %	1 %	1 %	1 %	1 %	1 %	1 %	1 %
Kommun	0 %	0 %	0 %	0 %	0 %	0 %	0 %	0 %	0 %
Ljusförhållande	8 %	7 %	6 %	5 %	5 %	3 %	2 %	2 %	2 %
Lokalpolisområde	0 %	0 %	0 %	0 %	0 %	0 %	0 %	0 %	0 %
Län	0 %	0 %	0 %	0 %	0 %	0 %	0 %	0 %	0 %
Månad	0 %	0 %	0 %	0 %	0 %	0 %	0 %	0 %	0 %
N-sweref	0 %	0 %	0 %	0 %	0 %	0 %	0 %	0 %	0 %
Okänt klockslag	99 %	99 %	99 %	99 %	99 %	99 %	99 %	99 %	99 %
Olycksplats	0 %	0 %	0 %	0 %	0 %	0 %	0 %	0 %	0 %
Olyckstyp	0 %	0 %	0 %	0 %	0 %	0 %	0 %	0 %	0 %
Olycksväg/-gata	5 %	6 %	5 %	6 %	6 %	5 %	9 %	7 %	7 %
Originalskiss saknas	0 %	0 %	0 %	0 %	0 %	0 %	0 %	0 %	0 %
Osäkert datum	100 %	100 %	100 %	100 %	100 %	100 %	100 %	100 %	100 %
Osäkert klockslag	99 %	99 %	99 %	99 %	99 %	99 %	100 %	100 %	100 %
Platstyp	0 %	0 %	0 %	0 %	0 %	0 %	0 %	0 %	0 %
Polisområde	0 %	0 %	0 %	0 %	0 %	0 %	0 %	0 %	0 %
Polisregion	0 %	0 %	0 %	0 %	0 %	0 %	0 %	0 %	0 %
Position i kartan	88 %	87 %	88 %	88 %	86 %	88 %	92 %	93 %	91 %
Rapport-ID	0 %	0 %	0 %	0 %	0 %	0 %	0 %	0 %	0 %
Registreringsnummer ID	17 %	18 %	20 %	20 %	22 %	20 %	22 %	19 %	20 %
Registreringstid (dagar)	0 %	0 %	0 %	0 %	0 %	0 %	0 %	0 %	0 %
Trafikanvisning väg A	33 %	34 %	31 %	30 %	29 %	23 %	16 %	16 %	14 %
Trafikanvisning väg B	88 %	89 %	87 %	88 %	87 %	87 %	84 %	84 %	83 %
Trafikreglering väg A	88 %	87 %	84 %	84 %	85 %	87 %	81 %	81 %	79 %
Trafikreglering väg B	99 %	99 %	99 %	99 %	98 %	95 %	92 %	92 %	92 %
Trafiksignal väg A	35 %	36 %	31 %	30 %	26 %	22 %	13 %	12 %	10 %
Trafiksignal väg B	86 %	87 %	85 %	86 %	84 %	86 %	83 %	83 %	82 %
Väderlek	12 %	12 %	10 %	10 %	9 %	10 %	9 %	9 %	8 %
Väglag	12 %	12 %	10 %	10 %	10 %	10 %	9 %	9 %	8 %
Vägtyp väg A	12 %	11 %	9 %	8 %	34 %	100 %	100 %	100 %	100 %
Vägtyp väg B	84 %	85 %	83 %	84 %	88 %	100 %	100 %	100 %	100 %
År	0 %	0 %	0 %	0 %	0 %	0 %	0 %	0 %	0 %

Tabell B1.2. Andel "Okänt", "Okänd", "Uppgift saknas", "Tom", "Okänt fordon", "Övrig/okänd plats" och "Okänd skadegrad" per personvariabel och år.

Variabel	2016	2017	2018	2019	2020	2021	2022	2023	2024
Antal släp	0 %	0 %	0 %	0 %	0 %	0 %	0 %	0 %	0 %
Backade	99 %	99 %	99 %	99 %	99 %	99 %	99 %	99 %	99 %
Fall ombord/från fordon	100 %	100 %	100 %	99 %	99 %	99 %	97 %	97 %	97 %
Fyrhjuling	100 %	100 %	100 %	100 %	71 %	1 %	0 %	0 %	0 %
Kön	6 %	6 %	6 %	7 %	7 %	7 %	6 %	6 %	6 %
Körde av vägen	84 %	85 %	86 %	87 %	86 %	85 %	85 %	85 %	85 %
Körde på fast föremål	92 %	93 %	93 %	93 %	92 %	91 %	91 %	91 %	91 %
Misstänkt alkoholpåverkad	27 %	30 %	32 %	32 %	33 %	28 %	26 %	25 %	26 %
Nationalitet	6 %	7 %	7 %	7 %	7 %	7 %	7 %	7 %	7 %
Omkörning	98 %	98 %	98 %	98 %	99 %	98 %	97 %	97 %	97 %
Skadegrad	8 %	8 %	8 %	8 %	9 %	9 %	7 %	7 %	7 %
Stod stilla/parkerat	100 %	100 %	100 %	100 %	100 %	99 %	99 %	99 %	99 %
Trafikantkategori - Huvudgrupp	0 %	0 %	0 %	0 %	0 %	0 %	0 %	0 %	0 %
Trafikantkategori - Undergrupp	0 %	0 %	0 %	0 %	0 %	0 %	0 %	0 %	0 %
Trafikanroll	4 %	4 %	3 %	3 %	2 %	3 %	2 %	2 %	2 %
U-sväng	100 %	100 %	100 %	100 %	100 %	100 %	100 %	99 %	100 %
Ålder	6 %	7 %	7 %	7 %	7 %	7 %	7 %	7 %	7 %
Övningskörning	100 %	100 %	100 %	100 %	99 %	99 %	100 %	100 %	100 %

Bilaga 2: Andel saknade värden sjukvård

Tabell B2.1. Andel "Tom", "Okänt", "Okänd", "Uppgift saknas", "Övrig/okänd plats", "Okänd skada" och "Värde mängd saknas" per variabel och år.

Variabel	2016	2017	2018	2019	2020	2021	2022	2023	2024
Alkoholpåverkan	80 %	80 %	81 %	82 %	71 %	47 %	78 %	90 %	90 %
Annan skyddsutrustning	75 %	76 %	77 %	79 %	54 %	7 %	6 %	6 %	6 %
Attribut till platstyp	94 %	94 %	94 %	93 %	92 %	92 %	92 %	92 %	92 %
Avliden datum	100 %	100 %	100 %	100 %	100 %	100 %	100 %	100 %	100 %
Barnstol/kudde	99 %	98 %	99 %	99 %	99 %	99 %	99 %	99 %	99 %
Bebyggelsestyp	3 %	3 %	3 %	4 %	4 %	12 %	17 %	17 %	15 %
Bälte	68 %	69 %	72 %	72 %	73 %	75 %	77 %	77 %	79 %
Dag	0 %	0 %	0 %	0 %	0 %	0 %	0 %	0 %	0 %
Drogpåverkan	100 %	100 %	100 %	100 %	83 %	51 %	84 %	96 %	96 %
Dödsorsak	100 %	100 %	100 %	100 %	100 %	100 %	100 %	100 %	100 %
E-sweref	5 %	5 %	5 %	5 %	6 %	15 %	21 %	21 %	21 %
Hjälm	69 %	70 %	69 %	66 %	57 %	65 %	65 %	67 %	66 %
Händelseförlopp	0 %	0 %	0 %	0 %	0 %	0 %	0 %	0 %	0 %
I Konflikt med - Huvudgrupp	60 %	62 %	63 %	63 %	63 %	68 %	68 %	68 %	70 %
I Konflikt med - Undergrupp	61 %	63 %	64 %	64 %	64 %	68 %	69 %	69 %	70 %
ISS	0 %	0 %	0 %	0 %	4 %	10 %	11 %	10 %	9 %
Inkom med	0 %	0 %	0 %	0 %	0 %	0 %	0 %	0 %	0 %
Klockslag	4 %	4 %	5 %	4 %	4 %	12 %	16 %	15 %	15 %
Kommun	0 %	0 %	0 %	0 %	0 %	2 %	3 %	4 %	4 %
Krockkudde utlöst	75 %	76 %	78 %	79 %	80 %	81 %	82 %	81 %	83 %
Kön	0 %	0 %	0 %	0 %	0 %	0 %	0 %	0 %	0 %
Län	0 %	0 %	0 %	0 %	0 %	1 %	1 %	1 %	1 %
MAIS	0 %	0 %	0 %	0 %	4 %	10 %	11 %	11 %	9 %
Mc-ställ	98 %	98 %	98 %	98 %	98 %	98 %	98 %	98 %	98 %
Månad	0 %	0 %	0 %	0 %	0 %	0 %	0 %	0 %	0 %
N-sweref	5 %	5 %	5 %	5 %	6 %	15 %	21 %	21 %	21 %
Nya ISS	1 %	0 %	1 %	0 %	4 %	10 %	11 %	10 %	9 %
Okänt datum	100 %	100 %	100 %	100 %	100 %	100 %	100 %	100 %	100 %
Okänt klockslag	96 %	96 %	95 %	96 %	96 %	88 %	84 %	85 %	85 %
Olyckan inträffade	6 %	6 %	6 %	6 %	6 %	13 %	20 %	19 %	18 %
Olycksplats	0 %	0 %	0 %	0 %	0 %	1 %	2 %	2 %	3 %
Olyckstyp	0 %	0 %	0 %	0 %	0 %	0 %	0 %	0 %	0 %
Olycksväg/-gata	21 %	22 %	23 %	23 %	21 %	26 %	32 %	32 %	32 %
Osäkert datum	100 %	100 %	100 %	100 %	100 %	100 %	100 %	100 %	100 %
Osäkert klockslag	98 %	98 %	99 %	98 %	98 %	95 %	93 %	94 %	95 %
Platstyp	1 %	1 %	1 %	1 %	1 %	6 %	10 %	9 %	8 %
Position i kartan	79 %	80 %	80 %	80 %	80 %	68 %	62 %	64 %	66 %
Rapport-ID	0 %	0 %	0 %	0 %	0 %	0 %	0 %	0 %	0 %
Registreringstid (antal dagar)	0 %	0 %	0 %	0 %	0 %	0 %	0 %	0 %	0 %
Sjukhus	0 %	0 %	0 %	0 %	0 %	0 %	0 %	0 %	0 %
Skadegrad	1 %	1 %	1 %	1 %	1 %	1 %	1 %	1 %	2 %

Variabel	2016	2017	2018	2019	2020	2021	2022	2023	2024
TS-journal insamlad via	100 %	82 %	0 %	0 %	0 %	0 %	0 %	0 %	0 %
Trafikantkategori - Huvudgrupp	0 %	0 %	0 %	0 %	0 %	0 %	0 %	0 %	0 %
Trafikantkategori - Undergrupp	0 %	0 %	0 %	0 %	0 %	0 %	0 %	0 %	0 %
Trafikantroll	0 %	0 %	0 %	0 %	0 %	1 %	1 %	2 %	2 %
Underlag till allvarligt skadade 1%	11 %	11 %	11 %	13 %	10 %	10 %	11 %	11 %	9 %
Underlag till allvarligt skadade 5%	11 %	11 %	11 %	13 %	10 %	10 %	11 %	11 %	9 %
Underlag till mycket allvarligt skadade 10%	11 %	11 %	11 %	13 %	10 %	10 %	11 %	11 %	9 %
Vägomständigheter	7 %	5 %	4 %	4 %	5 %	15 %	20 %	18 %	18 %
Vårdförlopp	0 %	0 %	0 %	0 %	0 %	0 %	0 %	0 %	0 %
Vårdtid (dagar)	87 %	88 %	88 %	89 %	87 %	87 %	87 %	87 %	87 %
Ålder	0 %	0 %	0 %	0 %	0 %	0 %	0 %	0 %	0 %
År	0 %	0 %	0 %	0 %	0 %	0 %	0 %	0 %	0 %

VTI, Statens väg- och transportforskningsinstitut, är ett internationellt framstående forskningsinstitut med uppdrag från regeringen att bedriva forskning och utveckling inom transportområdet, omfattande infrastruktur, trafik, transporter och trafikanter. Vi arbetar för att kunskapen om transportsektorn kontinuerligt ska förbättras och är på så sätt med och bidrar till att uppnå Sveriges transportpolitiska mål.

Verksamheten omfattar samtliga trafikslag och områdena väg- och banteknik, drift och underhåll, fordonsteknik, trafiksäkerhet, trafikanalys, människan i transportsystemet, miljö, planerings- och beslutsprocesser, transportekonomi samt transportsystem. Kunskapen som VTI tar fram ger beslutsunderlag till aktörer inom transportsektorn och tillämpas i många fall i såväl nationell som internationell transportpolitik.

VTI bedriver forskning på uppdrag i en tvärvetenskaplig organisation. Vi arbetar också med utredning och rådgivning, samt utför olika typer av tjänster inom mätning och provning. På institutet finns tekniskt avancerad forskningsutrustning av olika slag och körsimulatorer i världsklass. Dessutom finns ett laboratorium för vägmateriäl, ett mättekniskt laboratorium och ett krocksäkerhetslaboratorium.

Transportbiblioteket vid VTI är en nationell resurs för informationsförsörjning och informations spridning för alla trafikslag inom transportforskningsområdet.

I Sverige samverkar VTI med ledande universitet och högskolor som bedriver närliggande forskning och utbildning. Vi medverkar även kontinuerligt i internationella forskningsprojekt, framförallt i Europa, och deltar aktivt i internationella nätverk.

VTI är en uppdragsmyndighet som lyder under regeringen och hör till Landsbygds- och infrastrukturdepartementet. Vårt kvalitetsledningssystem är certifierat enligt ISO 9001 och vårt miljöledningssystem enligt ISO 14001. Vissa provningsmetoder vid våra laboratorier är dessutom ackrediterade.

Vi är omkring 250 medarbetare och finns i Linköping (huvudkontor), Stockholm, Göteborg och Lund.

vti

Statens väg- och transportforskningsinstitut • www.vti.se • vti@vti.se • +46 (0)13-20 40 00