

# Verdsetting basert på stordata og avslørte preferanser

## En vurdering av muligheter for analyser innenfor transport

TØI rapport 1882/2022 • Forfattere: Stefan Flügel, Christian Weber, Askill H. Halse, Ingunn O. Ellis • Oslo 2022 • 90 sider

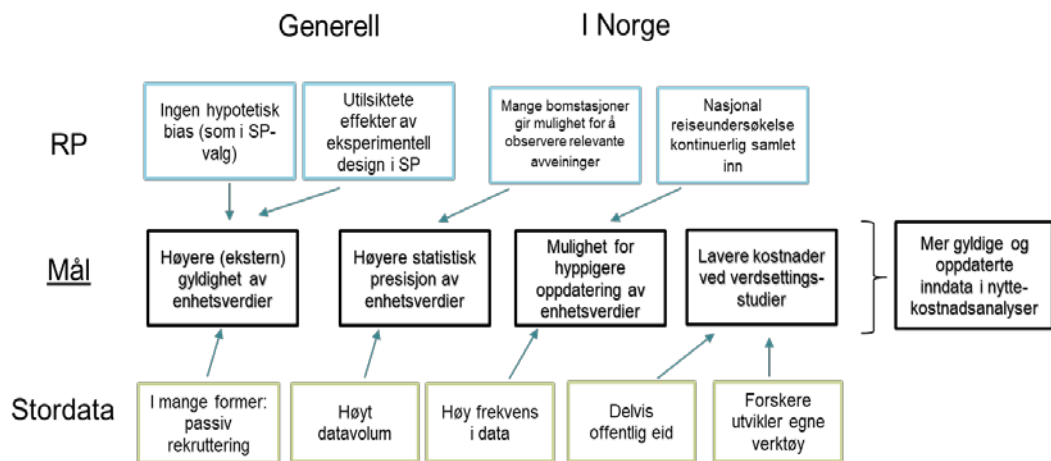
I flere tiår har uttalte («stated») preferanser (SP) vært den dominerende metoden for verdsetting innenfor transport. Det er imidlertid mye som tyder på at avslørte («revealed») preferanser (RP) gjør et sterkt comeback på grunn av tilgang til stordata og nye analysemuligheter som maskinlæring.

I denne rapporten vurderer vi muligheter som ligger i ulike RP-datakilder. Vi finner at data med GPS-sporing gir det bredeste og mest presise grunnlaget for verdsetting. For å kunne ivareta dagens segmentering av enhetsverdier, må man imidlertid samle inn ytterligere bakgrunnsdata. Bruken av tradisjonelle reisevaneundersøkelser er også rangert høyt, spesielt når synergier med estimering av transportmodeller kan realiseres.

### Bakgrunn og motivasjon

Mens SP-studier bygger på en analyse av hypotetiske valg i eksperimentelle omgivelser uten reelle konsekvenser for respondentene, blir RP-valg observert i virkelige omgivelser og er derfor den foretrukne metoden for å utlede preferanser. Men med RP-data har forskeren liten kontroll over dataene, og liten variasjon og/eller høy korrelasjon er en vedvarende utfordring i RP-basert estimering av enhetsverdier. Denne utfordringen kan delvis overvinnes med større datavolum, som blir mer og mer tilgjengelig på grunn av økningen av stordata.

Figur S1 oppsummerer hovedfordeler med RP-data generelt og med stordata, og illustrerer hvordan dette kan bidra til mer valide og mer oppdaterte enhetsverdier for analyser innenfor transport, deriblant nytte-kostnadsanalyser.



Figur S1: Oversikt over motivasjon for bruk av revealed preference (RP) data og stordata for transportrelatert verdsetting

## Arbeidsoppgaver og metode

Konklusjonene og anbefalingene i denne rapporten er basert på:

- 1) En litteraturgjennomgang om RP/stordata- basert verdsetting.
- 2) En liste over mulige datakilder og en diskusjon av deres relevans for verdsetting.
- 3) En vurdering av relevant kombinasjon av datakilder og enhetsverdier basert på 19 ulike kriterier. Poeng gis på en skala fra 1 til 5. Poengsummen er delvis basert på en intern Delphi-undersøkelse.
- 4) En syntese av vurderingen i tre grupper av kriterier: «Tilgang og generell kvalitet», «Analysemulighet for verdsettelse» og «Fleksibilitet, synergier og fremtidsperspektiv».
- 5) En praktisk beskrivelse av tre av de mest lovende tilnærmingene.
- 6) En casestudie for å illustrere noen utfordringer ved aggregerte datakilder.

Tabell S1 gir en oversikt over de inkluderte datakildene og noen sentrale egenskaper.

Tabell S1: Oversikt over datakilder.

Datakilde	Teknologi	(Antatt) eier av data / tilgang for forskerne	(Antatt) aggregeringsnivå	Mest egnet valgkontekst / enhetsverdi
Nasjonal RVU	Tradisjonell reisevaneundersøkelse	Transport- etatene / gratis	Disaggregert (enkeltreiser for personer)	Transportmiddelvalg / diverse
Mobildata	"Call Detail Record" via mobilmaster	Kommersielle tilbydere som Telia / kostbar	Aggregert (grunnkrets eller ruter)	Rutevalg (hovedsakelig lange distanser) / VTTS bil
App panel med GPS-sporing	GPS/A-GPS, GNSS	Forsker / fri tilgang til egne paneler	Disaggregert (enkeltreiser for personer)	Transportmiddel- og rutevalg / diverse
Trafikkelledata	Sensorer (typisk elektrisk induksjon)	NPRA / fri	Aggregert (punkter)	Rutevalg / VTTS bil
Bompasseringsdata	ANPR kamera and RFID («bombrikke»)	NPRA / fri (begrenset tilgang per i dag)	Disaggregert (biler over flere punkter)	Rutevalg / VTTS bil
Sporingsdata fra kommersielle tilbydere	Diverse (GPS, navigasjonsverktøy, ..)	Kommersielle tilbydere som TomTom eller Fitbit / kostbar	Aggregert (grunnkrets eller ruter)	Rutevalg / VTTS bil
Egeninstallerte kamera og sensorer	Diverse (ANPR, RFID, blåtann og magnetiske sensorer)	Forskere / fri tilgang til egeninstallerte maskinvare og data	Disaggregert (biler over flere punkter)	Rutevalg / VTTS bil
Mobility-as-a-Service bestillingsdata	Lagrete data fra apper	MaaS tilbydere som Bolt eller Ruter / uklar per i dag	Disaggregert (enkeltreiser for personer)	Diverse / VTTS (ventetid)
Passasjertellinger	Diverse (kamerateknologi, mobilsporing og/eller lysbarrierer)	Kollektivselskapene / fri (noen restriksjoner)	Aggregert (snitt/avgang)	Valg av driftsart/avgang / trengselsfaktorer
Kamerabaserte tellinger ved stasjoner	Kamera (underbygd av maskinlæring)	Forskere / fri tilgang til egeninstallerte maskinvare og data	Aggregert (stasjon/avgang)	Vente på neste avgang ved stasjon / trengselsfaktorer

## Sammendrag av evalueringen

Datatilgang og generell kvalitet ble vurdert basert på følgende kriterier:

- Tilgang til relevante og oppdaterte RP-data
- Ressurser som kreves for datatilgang og vedlikehold (høy poengsum for lave ressurser som kreves av det utførende organet for verdsettingsstudien; opprinnelige kostnader fra andre ikke inkludert)
- Ressurser som kreves for databehandling (høy poengsum for lave ressurser som kreves av det utførende organet for verdsettingsstudien; opprinnelige kostnader fra andre ikke inkludert)
- Datavolum
- Dekning (høy poengsum hvis hele Norge er dekket)
- Representativitet

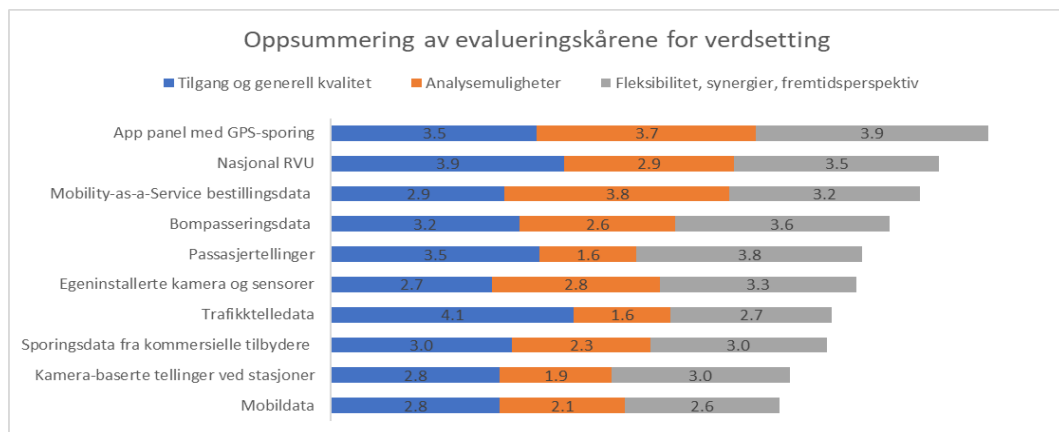
Mens de tre sistnevnte kriteriene kan avhenge av enhetsverdien, er den totale poengsummen for denne gruppen av kriterier ganske stabil på tvers av ulike relevante enhetsverdier av samme datakilde. Kriteriene for muligheter for analyse for verdsetting var:

- Observasjon av faktiske valg
- Kvantifisering av attributter og kostnader ved valgt alternativ
- Identifisering/modellering av ikke-valgte alternativer (valgsett)
- Kvantifisering av attributter og kostnader ved ikke-valgt alternativ
- Variasjon og korrelasjon i sentrale attributter
- Mulighet for kontroll for andre effekter
- Mulighet for segmentering (nåværende segmentering)
- Mulighet for kombinerte RP-SP modeller og andre avanserte estimeringsmetoder

Den siste gruppen av kriterier omfatter fleksibilitet, synergier og fremtidsperspektiv for datakildene. Denne gruppen vurderes ut fra et generelt perspektiv og ikke fra forskernes perspektiv (som de to foregående gruppene). Følgende kriterier var inkludert:

- Mulighet for hyppige og kontinuerlige datainnsamlinger i fremtiden
- Mulighet for å segmentere resultater utover dagens segmentering
- Synergier med transportmodeller
- Andre synergier
- Relevans for nye trender/teknologier

Figur S2 gir en samlet rangering av de evaluerte datatypene. Skårene for mulighet for analyse for verdsetting gjelder enhetsverdien med best score innenfor hver datatype.



Figur S2: Overordnet rangering av RP-datakilder for verdsetting.

App-panel med GPS-sporing er rangert høyest totalt sett.

Poengsummene for muligheter for analyse varierer med de underliggende enhetsverdiene.

En annen viktig informasjon er hvor mange enhetsverdier datakilden kan brukes for. Tabell S2 oppsummerer funnene våre.

Tabell S2: Antall relevante enhetsverdier og spenn i total skår for analysemulighet for estimering av enhetsverdier.

Datakilde	Antall relevante enhetsverdier *	Total skår	Hovedfordel	Hovedulempe
Nasjonal RVU	6	2,2- 2,9	Dekker dagens krav til segmentering	Unøyaktig stedsfesting
Mobildata	2	1,7-2,1	Noe bedre kontroll over ruter sammenlignet med telldata, i alle fall for lange reiser	Lite kontroll og muligheter for segmentering, fungerer dårlig for korte distanser
App panel med GPS-sporing	10	3,3-3,7	Detaljert informasjon om ruter	Reisehensikt upålitelig observert
Trafikktelldata	1	1,6		Ruter ikke direkte observert
Bompasseringsdata	2	2,6	Kan skille biltyper	Fungerer bare i nettverk med mange bomstasjoner
Sporingsdata fra kommersielle tilbydere	2	2,1-2,3	Bedre kontroll over ruten enn med telldata og mobildata	Lite bakgrunnsinformasjon
Egeninstallerte kamera og sensorer	4	2,7-2,9	Bra kontroll over ruter hvis god dekning av kameraene	Reisehensikt ikke observert
Mobility-as-a-Service bestillingsdata	2	3,5-3,8	Direkte og presis informasjon om attributtverdier	Reisehensikt ikke observert, bruk av app muligens endogent
Passasjertellinger	1	1,6		OD-relasjoner er ikke direkte observert
Kamerabaserte tellinger ved stasjoner	1	1,9		Fungerer kun under spesielle forhold

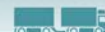
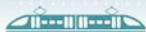
## Illustrasjons- og casestudie

Rapporten inneholder også en mer praktisk beskrivelse av tre av de mest lovende tilnærmingene (nasjonal RVU, Fotefar, som er en kommende GPS-app-sporingsprogramvare, og bompasseringsdata) samt en casestudie ved bruk av aggregerte datakilder (trafikktellinger, mobildata og data fra TomTom). Sistnevnte illustrerer noen av de praktiske vanskelighetene ved å bruke aggregerte data for å utlede enhetsverdier.

## Konklusjoner og anbefalinger

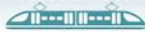
Nedenfor oppsummerer vi hovedkonklusjonene våre:

- 1) Per i dag er reiseundersøkelser som **nasjonal RVU** den mest relevante datakilden med tanke på dagens segmentering av enhetsverdier som krever informasjon om reiseformål. Det er store potensielle synergier med transportmodeller og man bør vurdere å samkjøre neste RTM-



estimering med neste verdsettingsstudie. I den forbindelse kan det være hensiktsmessig å gå bort fra dagens RVU, og heller utforme en mer skreddersydd undersøkelse som egner seg bedre både for etterspørselsmodellering og verdsetting.

- 2) Data fra apper som kan **spore individer med GPS** eller andre høyoppløselige og høyfrekvente sensorer scorer totalt sett best i vår vurdering. Mulighet for å legge til bakgrunnsinformasjon er viktig. Dette kan kreve ytterligere datainnsamling, for eksempel i form av spørreundersøkelser.
- 3) **En kombinasjon av spørreundersøkelser (og/eller registerdata) og GPS-sporing anses som det beste alternativet og noe som anbefales å jobbe mot.**
- 4) **Aggregerte data** (f.eks. trafikktelldata og passasjertelldata) legger store begrensninger på analysemuligheter og vil neppe være tilstrekkelig for nasjonale enhetsverdier gitt kravene til dekning og segmentering. Når det er sagt, kan det – basert på passende casestudier – bidra til å validere det absolutte nivået på tidsverdien (VTTS).
- 5) **Aggregerte mobildata** gir bedre analysemuligheter sammenlignet med telldata, i hvert fall for lange distanser, men er ganske dyrt å få tilgang til. Som andre aggregerte datakilder har mobildata klare begrensninger sammenlignet med mer disaggregerte datakilder.
- 6) **Bompaseringsdata** som sporer biler vil kunne gi informasjon om rutevalg for enkeltpersoner eller grupper i områder med god dekning av bomstasjoner. Det er ulike muligheter for å legge inn individuelle bakgrunnsvariabler. Slike data vil i de fleste tilfeller ikke være helt anonyme, men tilgang til ikke-anonyme data for forskningsformål vil mest sannsynlig være mulig under gjeldende personvernlover. Å lette tilgangen til data vil imidlertid kreve noe velvilje og innsats fra eierne av dataene. Et mer fleksibelt (men dyrere) alternativ til disse dataene er å sette opp **egne kameraer for automatisk nummerskiltgjenkjenning (ANPR)**.
- 7) **Aggregerte App-data fra kommersielle virksomheter** kan også være et lovende alternativ. Statens vegvesen har tilgang til aggregerte sporingsresultater fra f.eks. TomTom, en datakilde som kan brukes mer enn i dag for å studere ruteatferd, f.eks. ved bompenger over hele landet. For å bruke TomTom-data til forskning er tilgang til mer informasjon om datainnsamling og databehandling, og muligheten å dele denne informasjonen i offentlige rapporter/artikler er avgjørende. Det er for tiden også store begrensninger i å dele data og publisere resultater fra dataanalyse.
- 8) De fleste datakilder nevnt under 4) – 7) har en grunnleggende fordel i sin **passive rekruttering**. Datakildene er derfor interessante for kvalitets-sikring av undersøkelser og app-baserte studier der uobserverbare faktorer kan påvirke nivået på VTTS på grunn av utvalgsskjevhet. Når det er sagt, kan det også være noen skjevheter i utvalget hos mobilselskaper og app-dataleverandører.
- 9) En disaggregert datakilde med stort potensial er **MaaS-bestillingsdata** (f.eks. fra raid-hailing-tjenester). Slike data er for øyeblikket begrenset i tilgang og anvendelse. I Norge virker det mest aktuelt å studere valg/-



preferanser for mikromobilitet. Denne typen data kan også gjøres tilgjengelig via fremtidige versjoner av mer tradisjonelle kollektiv-apper (f.eks. via en fremtidig versjon av Ruter-appen som kan la reisende velge, bestille og betale for alle tilgjengelige transportløsninger).

Vi ser tre tilnærminger for neste verdsettingsstudie. Disse er gitt nedenfor i rangert rekkefølge.

- 1) **GPS-sporingsdata pluss bakgrunnsundersøkelser.** Rekrutteringen bør komme fra en kombinasjon av store (eksisterende) utvalg eller – helst – Folkeregisteret. Det bør gis økonomiske incentiver for å donere sporingsdata til prosjektet, da dette sannsynligvis vil tiltrekke seg et bredere utvalg og derfor kan redusere faren for utvalgsskjevheter. Fra et modelleringsperspektiv vil kombinerte transportmiddelvalg- og rutevalgmodeller sannsynligvis gi det beste og bredeste grunnlaget for estimering av enhetsverdier. Bakgrunnsundersøkelsen bør inkludere spørsmål om tilgang til transportmiddel, biltype og billettype og inkludere informasjon om stedsfesting av hjemmet, jobben og andre hyppige destinasjoner for respondentene slik at reisemålet kan utledes fra stedsfestingen i GPS-dataene. I tillegg kan små SP-eksperimenter inkluderes i bakgrunnsundersøkelsen for kryssvalidering og for estimering av enhetsverdier som kan være vanskelig å estimere basert på RP-data.
- 2) **Nasjonal RVU eller – fortrinnsvis – en skreddersydd reiseundersøkelse i felles estimering med RTM-modellen.** Sammenlignet med forslag (1.), legger denne tilnærmingen mindre vekt på presise data og legger vekt på konsistens og synergier med transportmodeller. Enhetsverdiene vil bli utledet fra nyttefunksjoner til transportmidlene i mode/destinasjonsvalgmodellen som er en del av RTM-systemet. Kalibrering av rutevalgmodeller i nettverksmodellen (f.eks. CUBE) mot aggregerte datakilder kan i tillegg støtte estimering/anbefaling av enhetsverdier. Det anbefales sterkt at romlig informasjon fra reiseundersøkelsesdataene gis gjennomgående med 8-sifrede grunnkretskoder (dvs. man bør gå bort fra gjeldende praksis med å gi grunnkretser med mindre enn 100 innbyggere 6-sifrede koder). Med det vil presisjonsnivået fortsatt være langt lavere enn ved GPS-sporing, men det bør være akseptabelt innenfor denne tilnærmingen.
- 3) En tredje tilnærming ville være å **beholde SP-metoden.** I dette tilfellet vil vi anbefale å bruke **flere godt utformede RP-casestudier for å validere/justere det overordnede nivået av VTTS.** Kombinerte RP-SP-modeller vil bli anbefalt for å utnytte fordelene med begge datatyper. I den forbindelse vil det være å foretrekke å rekruttere deler av SP-utvalget fra områdene hvor RP-casestudiene gjennomføres.