

Forschungsdatenbank Nichtwohngebäude
Abschlussstagung und Workshops
28. und 29. April 2021
(digital)

Das Stichprobenkonzept

Dr. Holger Cischinsky
Institut Wohnen und Umwelt
Darmstadt

1. Geodatenanalyse

- Bagatellbereinigung
- geometrische und semantische Attribute
- Flächendeckende, überlappungsfreie Erhebungsbezirke

2. Stichprobe

- 500 Erhebungsbezirke à 200 Hausumringe

3. Screening

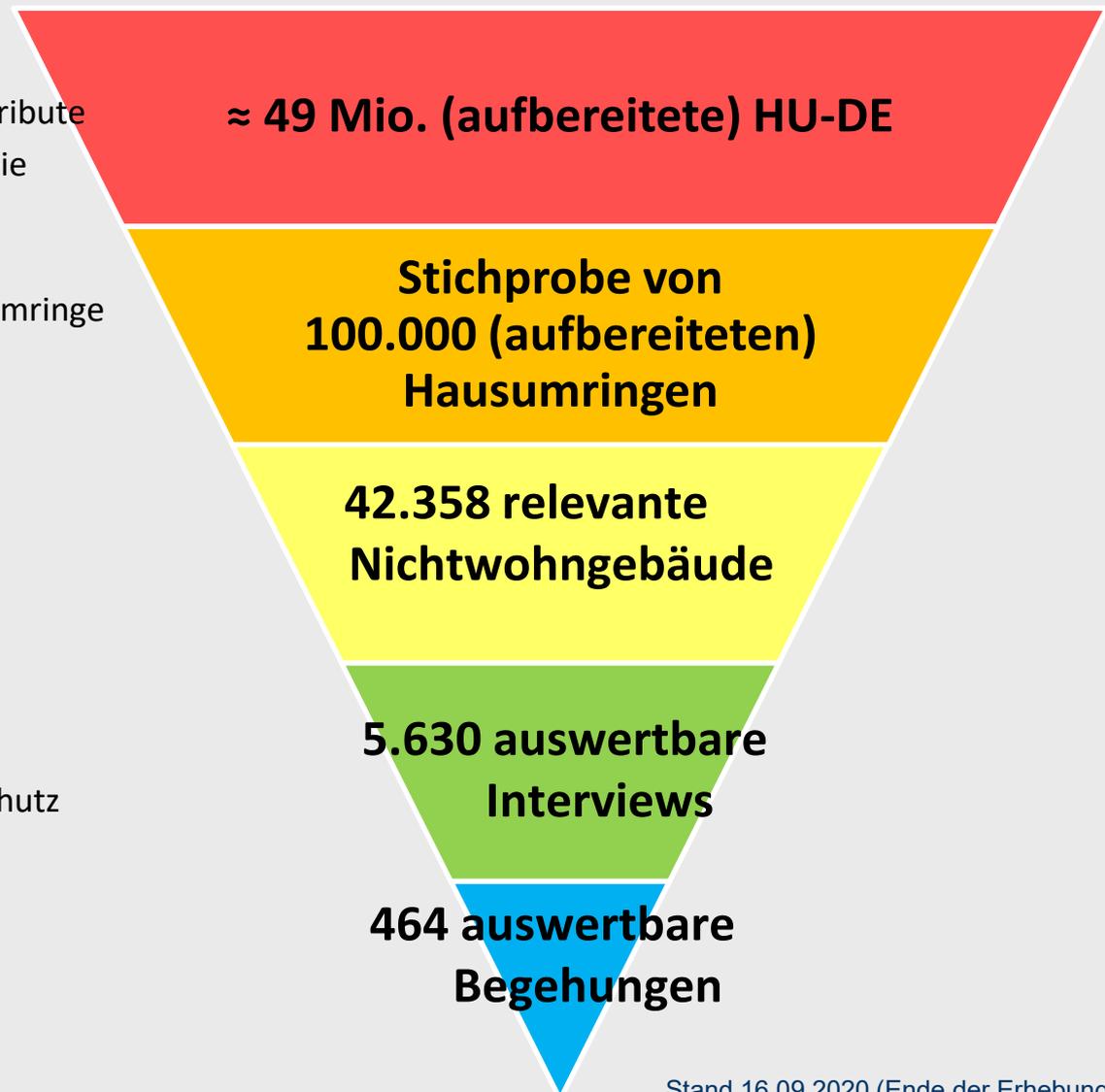
- Relevanz
- Beziehung Hausumring - Gebäude
- Adressdaten
- Gebäudebasismerkmale

4. Breitenerhebung

- Struktur-Merkmale
- energetische Qualität, Modernisierungsfortschritt und Modernisierungsrate bei Wärmeschutz und Gebäudetechnik
- Eigentümerstrukturen
- Bewirtschaftungsverhalten

5. Tieferhebung

- Energieverbrauch /-bedarf
- Nutzerparameter

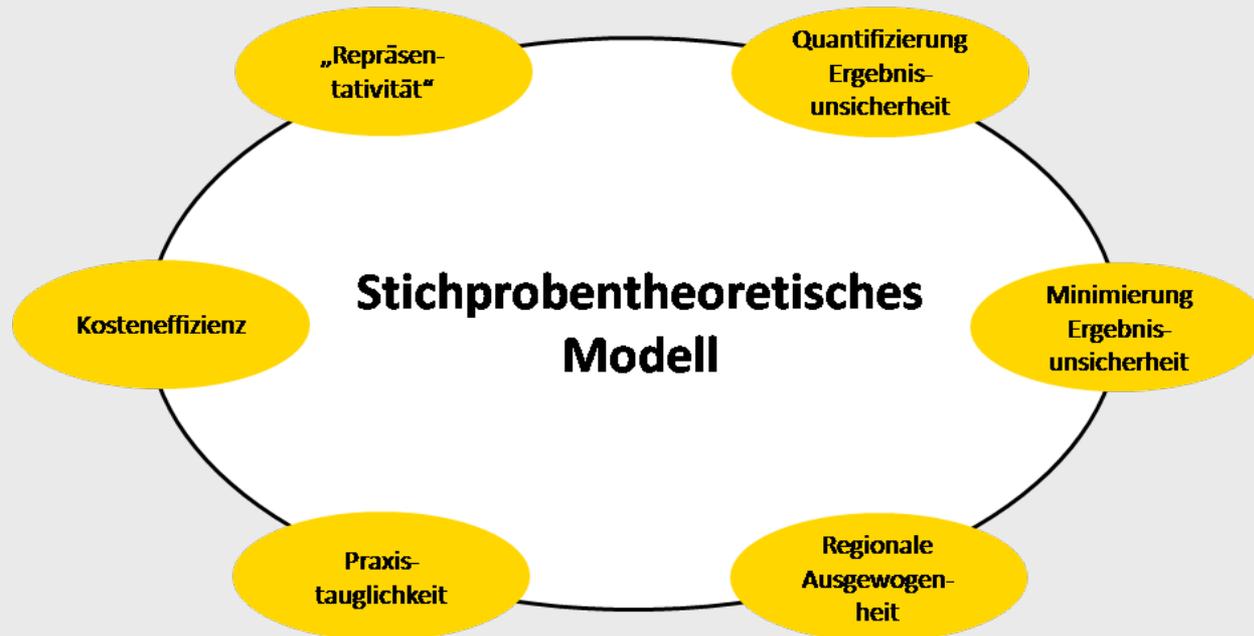


Stand 16.09.2020 (Ende der Erhebungen)

Aufgrund der **Unmöglichkeit einer Vollerhebung** aller Nichtwohngebäude Deutschlands benötigen wir ein stichprobentheoretisches Modell, das festlegt,

- **wie viele** Nichtwohngebäude stellvertretend in die Erhebung aufgenommen werden,
- **nach welchen Kriterien** diese Gebäude ausgewählt werden und
- **wie** die für diese Gebäude erhobenen Daten so **verdichtet** werden, dass hinreichend genaue Rückschlüsse über die interessierenden Verhältnisse und Strukturen in der Grundgesamtheit aller Nichtwohngebäude gezogen werden können.

Welche Anforderungen hatte das stichprobentheoretische Modell zu erfüllen?



„**Repräsentativität**“ im statistischen Sinne wird erreicht, wenn eine **Methodik** eingesetzt wird, die einen **unverzerrten** (d.h. erwartungstreuen) Rückschluss von der Stichprobe auf die Grundgesamtheit bei möglichst **geringer** und **quantifizierbarer stichprobenbedingter Unsicherheit** erlaubt.

→ „Repräsentativität“ ist keine Eigenschaft der gezogenen Gebädestichprobe, sondern ein **methodisches Gütekriterium**.

→ Alle drei Voraussetzungen für „Repräsentativität“ haben wir im Projekt erfüllt:

- **Zufallsprinzip**
- Kenntnis und Berücksichtigung von **Ziehungswahrscheinlichkeiten**
- **Positive Ziehungswahrscheinlichkeit** für jedes Nichtwohngebäude

Warum haben wir erst eine Stichprobe von Hausumringen und nicht gleich von Nichtwohngebäuden gezogen?

Ideale Auswahlgrundlage:

Liste aller Nichtwohngebäude in Deutschland (und nur dieser) mitsamt:

- Objektanschrift
- kompetenten Auskunftspersonen
- fundamentalen Gebäudeeigenschaften

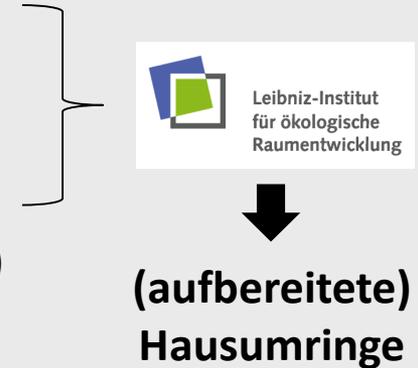
Objektanschrift	Ansprechpartner	Gebäudebeschreibung
Musterstraße 4 (Nr. 1) 12345 Musterhausen	Max Mustermann (Verwalter) Tel.: 0123456789 12345 Musterhausen	Bürogebäude 4 Vollgeschosse Baujahr 2000
Musterweg 17 54321 Musterdorf	Musterbau AG Frau Muster 99999 Musterstadt	Werkstatt 1 Vollgeschoss Baujahr 1960
Musterpfad 6 99999 Musterstadt	Autohaus Mustermann Herr Mustermann 99999 Musterstadt	Autohaus 2 Vollgeschosse Baujahr 2005
(...)	(...)	(...)

Abhilfe:

Datenbank amtlicher Hausumringe (**HU-DE**) (= 2D-Gebäudegrundriss (Teilauszug aus der amtlichen Liegenschaftskarte)) als **Auswahlgrundlage** für die Stichprobenziehung

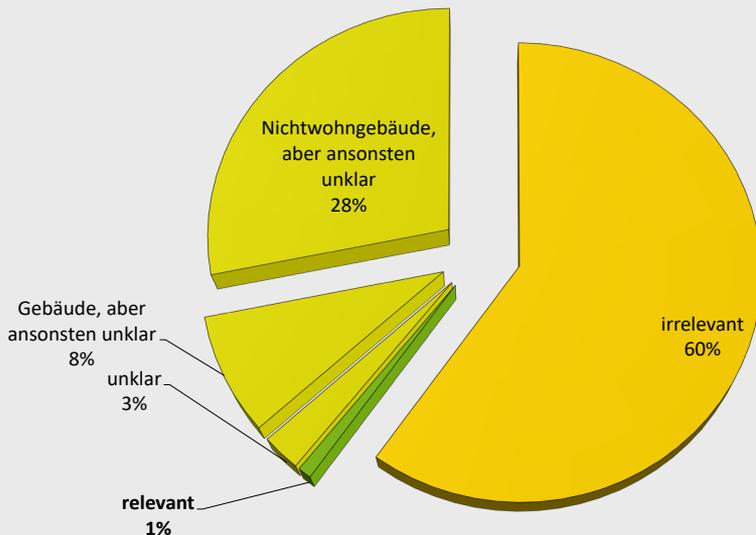
Welche „Probleme“ gingen mit der Nutzung der HU-DE-Datenbank als Auswahlgrundlage einher?

- Vorhandensein von Kleinstpolygonen, bei denen es sich nicht um eigenständige Gebäude handeln kann
- Topologische Inkonsistenzen (Überlappungen)
- Informationen zur treffsicheren Identifizierung von (GEG-relevanten) Nichtwohngebäuden teilweise nicht vorhanden
- Mutmaßlich größter Teil der HU-DE gehört nicht zu (GEG-relevanten) Nichtwohngebäuden („irrelevante Einheiten“)
- Adressinformationen teilweise fehlend
- Fehlende Informationen zu geeigneten Auskunftspersonen
- Nicht notwendigerweise 1:1-Äquivalenz zwischen HU-DE (Erhebungseinheit) und (Nichtwohn-) Gebäude (Untersuchungseinheit)



Wie sind wir damit umgegangen, dass die Relevanz vieler Hausumringe ex ante unbekannt ist?

Nur für einen Teil der knapp 49 Mio. Hausumringe der Auswahlgrundlage stand aufgrund **aussagekräftiger Nutzungsattribute** die Relevanz (Zugehörigkeit zu GEG-relevanten Nichtwohngebäuden) bzw. die fehlende Relevanz bereits ex ante relativ sicher fest. Häufig war die **Relevanz vieler Hausumringe** jedoch **unsicher** oder sogar gänzlich **unbekannt**.



Vom Projektteam auf Basis verfügbarer Nutzungsattribute vorgenommene Relevanzeinstufung der Hausumringe

Achtung: Die Einstufung als relevant oder irrelevant ist mit einer Restunsicherheit verbunden !!!

→ „Einfache“ Zufallsstichprobe von Hausumringen liefert erwartungsgemäß zu viel „Ausschuss“.

Lösung: Vornahme einer **Relevanzeinstufung** und Ziehung der Hausumring-Stichprobe so, dass relevante Hausumringe eher in die Stichprobe gelangen als irrelevante

Wie sind wir damit umgegangen, dass die Relevanz vieler Hausumringe ex ante unbekannt ist?

Den rund 19 Mio. Hausumringen ohne Informationen zur Untersuchungsrelevanz (GEG-relevantes Nichtwohngebäude) werden über **binär-logistischen Regressionsansatz** sog. **Relevanzwahrscheinlichkeiten** (Wahrscheinlichkeiten für Untersuchungsrelevanz) zugespielt.

Relevanz	Merkmal 1	Merkmal 2	Merkmal 3	(...)	prob (Relevanz)
1	a	908	1,8		1,00
0	b	160	3,1		0,00
0	c	253	2,3		0,00
0	c	91	5,2		0,00
1	a	476	1,2		1,00
1	b	806	1,9		1,00
0	c	321	5,3		0,00
0	c	102	3,9		0,00
0	b	387	4,1		0,00
	b	163	1,8		0,18
	c	240	3,9		0,08
	c	175	3,1		0,09
	a	614	3,4		0,82
	a	490	1,4		0,98
	c	361	4,8		0,62
	b	198	3,1		0,32
	a	956	1,9		0,91

Binär-logistische
Regressionsgleichung

Summe = 4,0, d.h. im Erwartungswert
sind 4 der 8 Hausumringe relevant.

→ Gruppierung der Hausumringe entsprechend ihrer Relevanzwahrscheinlichkeiten in Schichten und **Ziehung einer disproportional geschichteten Stichprobe**

→ Auch Hausumringe mit einer (unterstellten) Relevanzwahrscheinlichkeit von 0 hatten eine (geringe) Ziehungschance !!!

Warum war eine Screeningphase mit 100.000 Stichproben-Hausumringen erforderlich?

- Einholen von **Informationen zu potentiellen Auskunftspersonen**
- **Relevanzfeststellung** in Bezug auf die Stichproben-Hausumringe
 - ▶ Identifizierung von 42.358 breitenerhebungsrelevanten Gebäuden, bei denen es sich – mutmaßlich – um GEG-relevante Nichtwohngebäude handelt
- Abklärung des konkreten **Beziehungsgeflechts** zwischen Hausumring und Nichtwohngebäude
- **Erhebung weiterer Gebäudebasismerkmale** vor Ort
 - ▶ Auf sehr breiter Stichprobenbasis (100.000 Hausumringe) lassen sich **wesentliche Strukturen der Grundgesamtheit aller deutscher Nichtwohngebäude** aufdecken.
 - ▶ Ermöglichung der Identifizierung und **Korrektur etwaiger systematischer Gebäudeausfälle** bei der nachgelagerten Breitenerhebung
 - ▶ Bereitstellung von **Testdaten** zur (Weiter-) Entwicklung von Algorithmen zum geoinformatrischen Identifizieren von Nichtwohngebäuden



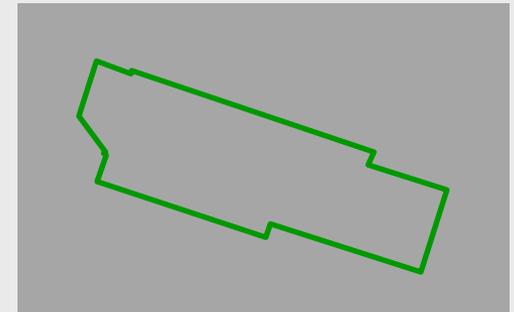
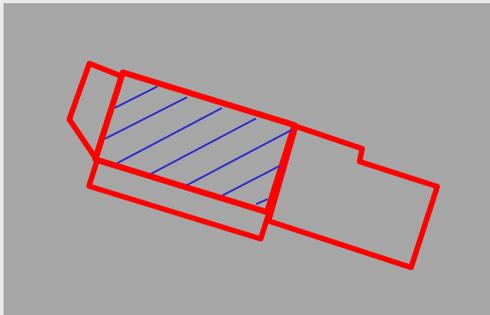
Wie haben wir hausumringspezifische in gebäude-spezifische Ziehungswahrscheinlichkeiten überführt?

Zwischen Hausumring und Gebäude besteht nicht notwendigerweise eine 1:1-Beziehung !!!

„Hausumring-Welt“
(Erhebungseinheiten)

Screening

„Gebäude-Welt“
(Untersuchungseinheiten)



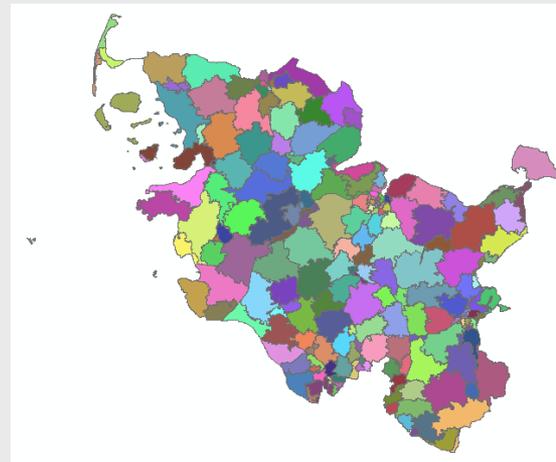
Lösung: Entwicklung einer geeigneten, für alle denkbaren Hausumring-Gebäude-Beziehungsgeflechte anwendbare **Formel**, durch die Ziehungswahrscheinlichkeiten von Hausumringen in Ziehungswahrscheinlichkeiten von Gebäuden umgerechnet werden können

Warum gab es Erhebungsbezirke und wie groß sind diese?

Die **Relevanzfeststellung** während des Screenings, das Einholen von **Informationen zu potentiellen Auskunftspersonen**, die Abklärung des konkreten **Beziehungsgeflechts** zwischen Hausumring und Nichtwohngebäude und die **Aufzeichnung von Gebäudebasismerkmalen** bedingen eine **Inaugenscheinnahme** der in die Bruttostichprobe aufgenommenen 100.000 Hausumringe **vor Ort**. Die Vorauswahl von Erhebungsbezirken **reduziert die damit einhergehenden Wegekosten**.

→ Zuschnitt der Erhebungsbezirke unter Rückgriff auf Relevanzwahrscheinlichkeiten und unter Beachtung diverser methodischer Vorgaben dergestalt, dass die Erhebungsbezirke **erwartungsgemäß ähnlich große Nichtwohngebäude-Bestände** haben

Beispiel: Erhebungsbezirke in Schleswig-Holstein:



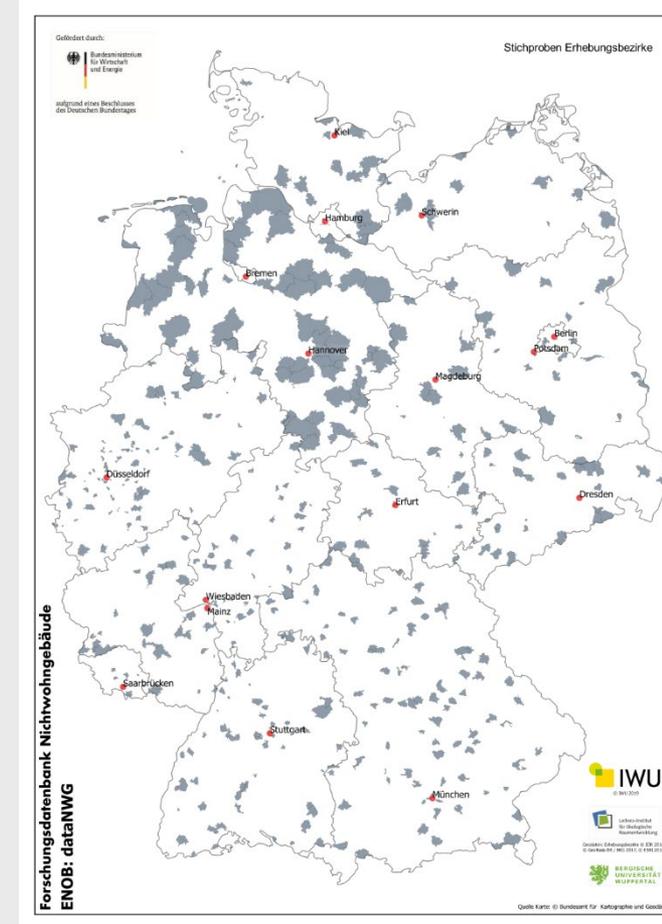
Wie erklärt sich die Festlegung auf 500 Erhebungsbezirke und wie wurden diese ausgewählt?

Festlegung auf 500 Stichproben-Erhebungsbezirke

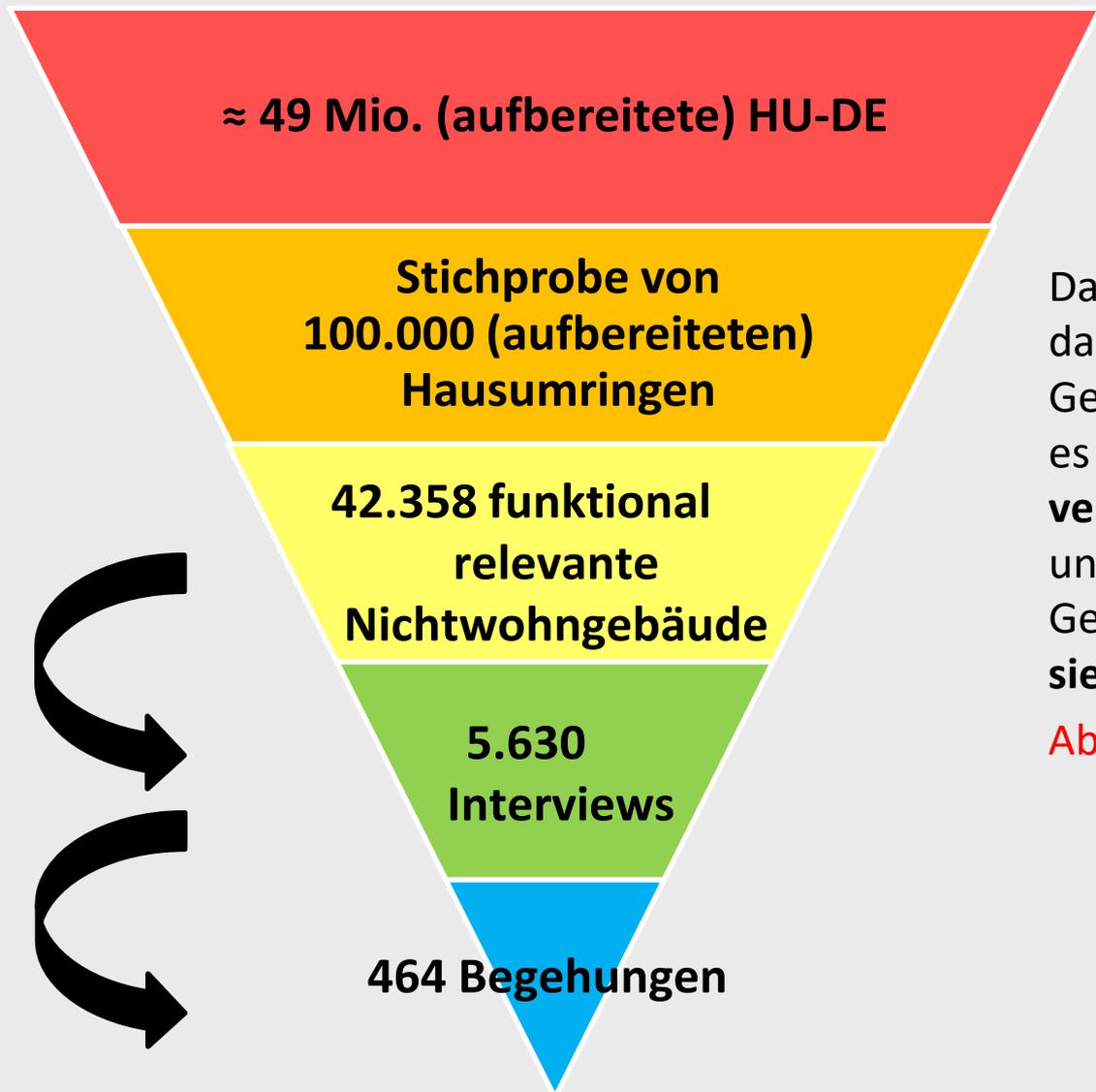
→ Kompromiss aus zwei entgegengesetzten Anforderungen:

- Ausreichend große Zahl, um sog. **Klumpeneffekte einzudämmen**, die sich negativ auf die Ergebnisschärfe auswirken, und um eine **befriedigende deutschlandweite Flächenabdeckung** sicherzustellen
- 500 Erhebungsbezirke **organisatorisch noch beherrschbar**

Die Auswahl der Erhebungsbezirke erfolgte **zufalls-gesteuert** und **geschichtet nach Landesteilen (Bundesland und Raumordnungsregion)**. Die Schichtung gewährleistet ex ante, dass am Ende aus allen Teilen Deutschlands Erhebungsdaten vorliegen.



Wie sind wir mit Gebäudeausfällen (Unit Nonresponse) während der Breiten- bzw. Tieferhebung umgegangen?



Da auch für „ausgefallene“ Gebäude dank vorausgegangenem Screening Gebäudebasismerkmale vorliegen, war es möglich, **systematische und damit verzerrende Ausfälle zu identifizieren** und mittels geeigneter statistischer Gewichtungsverfahren **zu kompensieren**.

Aber: Diese Korrektur erfolgte **nur** für die Breiten-, **nicht** für die Tieferhebung !!!

Sind die hochgerechneten Daten aus dem Screening oder die aus der Breitenerhebung „besser“?

Screening

- ⊕ Sehr großer Stichprobenumfang
- ⊕ Kein Potential für verzerrenden Nonresponse vorhanden
- ⊖ Begrenzter Merkmalskatalog
- ⊖ Mutmaßlich eingeschränkte Validität der erhobenen Merkmale

Breitenerhebung

- ⊖ Begrenzter (Netto-) Stichprobenumfang
- ⊖ Potential für verzerrenden Nonresponse nicht vollständig zu beseitigen
- ⊕ Umfangreicher Merkmalskatalog
- ⊕ Von hoher Validität der erhobenen Merkmale auszugehen

→ Frage, welche Hochrechnungsergebnisse besser sind bzw. – allgemein – welche Datenerhebungsmethode besser ist, lässt sich **pauschal nicht beantworten !!!**

Wo finde ich vertiefte Informationen zum Stichprobenkonzept bzw. zum methodischen Vorgehen?

ZWISCHENBERICHT

ENOB: dataNWG
Forschungsdatenbank Nichtwohngebäude
Förderkennzeichen 03 ET1315 A

Institut Wohnen und Umwelt GmbH
Forschungseinrichtung des Landes Hessen und der Stadt Darmstadt
Rheinstraße 65
64295 Darmstadt
Germany
Tel: +49 (0)6151 / 2904-0
Fax: +49 (0)6151 / 2904-97
Info@iwu.de
www.iwu.de

Leibniz-Institut für ökologische Raumentwicklung

BERGISCHE UNIVERSITÄT WUPPERTAL

Gefördert vom
Bundesministerium für Wirtschaft und Energie
im Förderbereich
Energieoptimierte Gebäude und Quartiere
der Fördermaßnahme *Anwendungsorientierte nichtnukleare FuE* des 6. Energieforschungsprogramms der Bundesregierung

10. Februar 2021

Inhalt

1	Zum Aufbau des Berichts	1
2	Anforderungen an das stichprobentheoretische Modell	3
3	Stichprobendesign und Schätzung	6
3.1	Auswahlgrundlage	6
3.1.1	Defizit der Auswahlgrundlage: Fehlende Informationen zu Gebäudeansprechpartnern	9
3.1.2	Defizit der Auswahlgrundlage: Fehlende Übereinstimmung zwischen Hausumringen und Gebäuden	10
3.1.3	Defizit der Auswahlgrundlage: Existenz irrelevanter Hausumringe	17
3.1.4	Defizit der Auswahlgrundlage: Untererfassung	25
3.1.5	Defizit der Auswahlgrundlage: Fehlende Gebäudemerkmale	28
3.2	Stichprobenauswahlverfahren	29
3.2.1	Stichprobenumfang	29
3.2.2	1. Ziehungsstufe: Geschichtete Zufallsauswahl von Erhebungsbezirken	32
3.2.3	2. Ziehungsstufe: Geschichtete Zufallsauswahl von Hausumringen	41
3.2.4	Ziehungswahrscheinlichkeiten	47
3.3	Schätzung	55
3.3.1	Schätzung von Merkmalssummen	55
3.3.2	Schätzung des Verhältnisses zweier Merkmalssummen	57
3.3.3	Kombinierte Schätzung von Merkmalssummen und Gaußsches Fehlerfortpflanzungsgesetz	59
4	Nicht-stichprobenbedingte Fehler	63
4.1	Fehler infolge von Nichtbeobachtbarkeit	63
4.1.1	Undercoverage	63
4.1.2	Nonresponse	64
4.2	Beobachtungsfehler	74
4.2.1	Messfehler	74
4.2.2	Datenaufbereitungsfehler	75
5	Literatur	76

ENOB:dataNWG

i

Forschungsdatenbank Nichtwohngebäude (ENOB:dataNWG)

Wir haben den
Bestand der Nichtwohngebäude
vermessen.

